Lecture Notes on Linear Algebra

Ornela Mulita
Master Degree in DSSC - Data Science & Scientific Computing
University of Trieste

September 2020

Preface

This text contains the lecture notes for an introductory course on linear algebra given in Falls, 2018-2020 to the students of the Master Degree in Data Science & Scientific Computing (DSSC) at the University of Trieste.

They correspond to a first course on linear algebra, which does not rely on any prerequisite, and can be considered as preceding more advanced numerical linear algebra topics in the succeeding courses on DSSC. All necessary notions are introduced from scratch, and several examples and exercises are provided with detailed explanations.

The main scope of the course is to emphasize the basic concepts of linear algebra (like vector spaces, matrices, linear transformations, scalar products, norms, eigenpairs etc.) as mathematical structures that are used to model the world around us. Once "persuaded" of this truth, students can perform computations in their future studies, while at the same time understand both the abstract and concrete concepts behind them.

The material in these notes is the result of careful elaboration and adaption from several classical books on linear algebra. The main sources of inspiration have been the books of S. Lang [1], G.H. Golub and C.F. Van Loan [2], D.C. Lay, S.R. Lay and J.J. McDonald [3], C.D. Meyer [4], G. Strang [5], A. Quarteroni, R. Sacco and F. Saleri [6].

Comprehensibly, it is impossible to provide in-depth coverage of all arguments and examples. Students who wish a more detailed coverage are warmly invited to consul the referenced books, which provide broad coverage of the field.

The notes correspond pretty closely to what I said in class in each academic year. I am thankful to the student A. Mecchina for partial typing in LaTeX.

Contents

1 Introduction **7**

2 Vector spaces **9**
2.1 Vector spaces and subspaces. 9
2.2 System of generators, linear dependence and independence of vectors. 14
2.3 Bases and dimension of a vector space. 15
2.4 Exercises 17

3 Matrices **19**
3.1 The space of matrices 19
3.2 Matrix operations 21
3.3 Matrix multiplication 22
3.4 Matrix inversion 25
3.5 Matrix transposition 26
3.6 Determinants 28

4 Linear mappings **33**
4.1 The space of linear mappings 33
4.2 Image and Kernel of linear mappings 35
4.3 Linear mappings and matrices 36
4.4 Exercises 39

5 Scalar products, norms and orthogonality **41**
5.1 Scalar products on vector spaces 41
5.2 Norms on vector spaces 44
5.3 Matrix norms 48
5.4 Exercices 51

6 Eigenvalues and eigenvectors **53**
6.1 Eigenvalues and eigenvectors 53
6.2 Computation of eigenpairs 55

6.3 Diagonalization of a matrix 59

Chapter 1

Introduction

Linear algebra is central to many areas of mathematics. Its influence ranges from algebra, through geometry, to applied mathematics. Indeed, some of the oldest and most widespread real-world applications of mathematics derive from linear algebra. Nowadays, the field of applications that use the language and tools developed in linear algebra has widened steadily and encompasses a vast diversity of problems in multidimensional calculus, differential geometry, functional analysis, multivariate statistics, control theory, dynamical systems, optimization, linear programming, computer graphics, genetics, graph theory, cryptography etc. More generally, linear algebra is used in most sciences and engineering areas, because it allows modeling many real-world phenomena, and efficiently computing with such models. Specially, linear algebra is the mathematics of data science: matrices and vectors are the language of data.

The first part of this course aims an introduction to both the theory (vector spaces and linear transformations between them) and the practice (matrices), and most importantly, the connection between these. Indeed, at the heart of the subject is the understanding and realization that, on finite-dimensional spaces, the theory of linear transformations is essentially the same as the theory of matrices. This is due primarily to the fundamental fact that the action of a linear transformation on a vector is exactly matrix multiplication between the coordinates of the transformation and the coordinates of the vector. Therefore linear transformations on finite-dimensional spaces will always have matrix representations, and vice versa.

Very often, in order to quantify errors we need to compute the magnitude of a vector or a matrix. For that purpose we will study the concepts of vector and matrix norms, followed by the concepts of inner products and orthogonality. In fact, many geometric structures (lengths, angles, etc.) can all be derived from inner products. These concepts provide powerful geometric tools for solving many engineering and applied sciences problems. In particular, they play a crucial role in the analysis of matrix algorithms. For instance, they are useful for assessing the accuracy of computations or for measuring progress during an iteration. Lastly, we will study the fundamental notions of eigenvalues and eigenvectors, with countless applications in real-world problems.

Chapter 2

Vector spaces

2.1 Vector spaces and subspaces.

Definition 2.1 (Field)**.** A **field** $\mathbb{K}$ is an algebraic structure consisting of a non-empty set of objects –called **scalars** (or **numbers)**– in which two operations are defined

- **addition** "+": $\forall a, b \in \mathbb{K} \implies a + b \in \mathbb{K}$;
- **multiplication** "$\cdot$": $\forall a, b \in \mathbb{K} \implies a \cdot b \in \mathbb{K}$ (often written as ab);

and the following properties are satisfied:

(F1) "+" is commutative: $\forall a, b \in \mathbb{K} \implies a + b = b + a$;

(F2) "+" is associative: $\forall a, b, c \in \mathbb{K} \implies (a + b) + c = a + (b + c)$;

(F3) there exists an element $0_{\mathbb{K}}$ (called **zero** or the **null element**) such that $\forall a \in \mathbb{K} \implies a + 0_{\mathbb{K}} = 0_{\mathbb{K}} + a = a$ (existence of the additive identity);

(F4) $\forall a \in \mathbb{K}$ there exists its **opposite** $-a \in \mathbb{K}$ such that $a+(-a) = (-a)+a = 0_{\mathbb{K}}$ (existence of the additive inverse);

(F5) "$\cdot$" is commutative: $\forall a, b \in \mathbb{K} \implies ab = ba$;

(F6) "$\cdot$" is associative: $\forall a, b, c \in \mathbb{K} \implies (a \cdot b)c = a(b \cdot c)$;

(F7) there exists an element $1_{\mathbb{K}} \neq 0_{\mathbb{K}}$ such that $\forall a \in \mathbb{K}$, $1_{\mathbb{K}} \cdot a = a$ (existence of the multiplicative identity);

(F8) $\forall a \neq 0_{\mathbb{K}} \in \mathbb{K}$, there exists its **inverse** $a^{-1} \in \mathbb{K}$ such that $a^{-1}a = 1_{\mathbb{K}}$ (existence of the multiplicative inverse);

(F9) the distributive property: $\forall a, b, c \in \mathbb{K} \implies (a+b)c = ac+bc$ and $a(b+c) = ab + ac$.

Remark 2.1. Properties (F1) − (F9) are also called **axioms** of a field. Using only these axioms, one can easily deduce that the additive and multiplicative identity in $\mathbb{F}$ are unique, the additive inverse (opposite) of an element of $\mathbb{F}$ is unique, and the multiplicative inverse (or, simply, inverse) of a nonzero element of $\mathbb{F}$ is unique.

Remark 2.2. We shall equivalently say that the field $\mathbb{K}$ is closed under the operations of addition and multiplication. The essential thing about a field is that it is a set of numbers which can be added and multiplied, where the ordinary rules of arithmetic hold true and where one can divide by nonzero elements.

Example 2.1 (Examples of fields)**.** Important examples of fields are the field of rational numbers $\mathbb{Q}$, the field of real numbers $\mathbb{R}$, the field of complex numbers $\mathbb{C}$, and the prime field with p elements $\mathbb{Z}_p$ (sometimes also written as $\mathbb{F}_p$), where p is any prime number. The set $\mathbb{Z}$ of integers is not a field. Indeed, in $\mathbb{Z}$ axioms (F1)-(F7) and axiom (F9) all hold, but axiom (F8) does not: the only nonzero integers that have multiplicative inverses are 1 and −1. For instance, $7 \neq 0_{\mathbb{Z}} \in \mathbb{Z}$ does **not** have a multiplicative inverse 7^{-1} in $\mathbb{Z}$ such that $7 \cdot 7^{-1} = 1_{\mathbb{Z}}$, since $7^{-1} \notin \mathbb{Z}$. Similarly, the set $\mathbb{N}$ of natural numbers is not a field either: not only nonzero naturals don't have their multiplicative inverses in $\mathbb{N}$, but also any natural number doesn't have its opposite in $\mathbb{N}$.

Definition 2.2 (Vector space)**.** A **vector space** V over the numerical field $\mathbb{K}$ is an algebraic structure consisting of a non-empty set of elements –called **vectors**– in which two operations are defined

- **addition** "+": $\forall \mathbf{v}_1, \mathbf{v}_2 \in V \implies \mathbf{v}_1 + \mathbf{v}_2 \in V$;
- **multiplication by scalars** "·": $\forall a \in \mathbb{K}, \forall \mathbf{v} \in V \implies a \cdot \mathbf{v} \in V$ (often written as $a\mathbf{v}$);

and the following properties are satisfied:

(V1) "+" is commutative: $\forall \mathbf{v}_1, \mathbf{v}_2 \in V \implies \mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$;

(V2) "+" is associative: $\forall \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in V \implies (\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$;

(V3) there exists an element $\mathbf{0}_V \in V$ (called **zero** or the **null vector**) such that $\forall\, \mathbf{v} \in V \implies \mathbf{v} + \mathbf{0}_V = \mathbf{v}$;

(V4) $\forall \mathbf{v} \in V$ there exists its **opposite** $(-\mathbf{v}) \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}_V$;

(V5) $\forall \mathbf{v} \in V \implies 1_{\mathbb{K}} \cdot \mathbf{v} = \mathbf{v}$;

(V6) the distributive property: $\forall a \in \mathbb{K}, \forall \mathbf{v}_1, \mathbf{v}_2 \in V \implies a(\mathbf{v}_1 + \mathbf{v}_2) = a\mathbf{v}_1 + a\mathbf{v}_2$;

(V7) the distributive property: $\forall a, b \in \mathbb{K}, \forall \mathbf{v} \in V \implies (a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$;

(V8) the associative property: $\forall a, b \in \mathbb{K}, \forall \mathbf{v} \in V \implies (ab)\mathbf{v} = a(b\mathbf{v})$.

Remark 2.3. Properties (V1) – (V8) are also called **axioms** of a vector space. Using only these axioms, one can easily show that the zero vector in Axiom (V3) is unique, as well as the vector $(-\mathbf{v})$ in Axiom (V4) is unique for each $\mathbf{v} \in V$.

Remark 2.4. We shall equivalently say that V is closed under vector addition and multiplication by scalars.

Remark 2.5. If $\mathbb{K} = \mathbb{R}$, V is called a **real vector space** and if $\mathbb{K} = \mathbb{C}$, V is called a **complex vector space**. Henceforth, we shall restrict our attention to vector spaces defined over the fields of real and complex numbers.

Example 2.2 (Examples of vector spaces)**.** Let us consider some remarkable examples of vector spaces.

1. $V = \{\mathbf{0}_V\}$.

2. Let $\mathbb{K}$ be a field. Then $\mathbb{K}$ is a vector space over itself. Addition "+" is addition between scalars $a + b$ and multiplication by scalars "$\cdot$" is multiplication between scalars ab, with $a, b \in \mathbb{K}$.

3. The spaces $\mathbb{K}^n$, for $n \geq 1$, of n-tuples of real (or complex) numbers are premier examples of vector spaces. Let

$$\mathbf{v} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad \mathbf{w} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \tag{2.1}$$

be elements of $\mathbb{K}^n$. The scalars $x_1, \ldots, x_n$ are called the **components**, or **coordinates**, of $\mathbf{v}$. Then addition and multiplication are defined coordinatewise, i.e., for $\mathbf{v}, \mathbf{w} \in \mathbb{K}^n$ and $a \in \mathbb{K}$ we have that

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix} \quad \text{and} \quad a\mathbf{v} = \begin{bmatrix} ax_1 \\ \vdots \\ ax_n \end{bmatrix}. \tag{2.2}$$

The zero element is the n-tuple with all its coordinates equal to 0. One can easily verify that properties (V1)-(V8) are satisfied. The most familiar examples are $\mathbb{R}^2$ and $\mathbb{R}^3$, which we can think of geometrically as the points of the ordinary two- and three-dimensional space, equipped with a coordinate system.

4. An example of vector space that arises in many applications is given by the vector space of functions. Let $X(\mathbb{D})$ be the set of all real-valued functions defined on a set $\mathbb{D} \subseteq \mathbb{R}$ in the variable t

$$X(\mathbb{D}) = \{\mathbf{f} : \mathbb{D} \to \mathbb{R},\, t \mapsto \mathbf{f}(t)\}. \tag{2.3}$$

Typically, $\mathbb{D}$ is the set of real numbers $\mathbb{R}$ or some interval $[a, b]$ on the real line. Given $\mathbf{f}, \mathbf{g} \in X(\mathbb{D})$, their addition $\mathbf{f} + \mathbf{g} \in X(\mathbb{D})$ is defined by

$$(\mathbf{f} + \mathbf{g})(t) := \mathbf{f}(t) + \mathbf{g}(t). \tag{2.4}$$

The scalar multiple $(a\mathbf{f})$, $a \in \mathbb{R}$ is the function defined by

$$(a\mathbf{f})(t) := a\mathbf{f}(t). \tag{2.5}$$

For instance, if $\mathbb{D} = \mathbb{R}$, $\mathbf{f}(t) = 2 + \sin 3t$, and $\mathbf{g}(t) = -1 + 1.5t$, then

$$(\mathbf{f} + \mathbf{g})(t) = 1 + \sin 3t + 1.5t \quad \text{and} \quad (2\mathbf{g})(t) = -2 + 3t. \tag{2.6}$$

We shall say that two functions in $X(\mathbb{D})$ are equal if and only if their values are equal for every t in $\mathbb{D}$. The zero function is the function that is *identically* zero, that is $\mathbf{0}(t) = 0$ for all $t \in \mathbb{D}$. Clearly, it acts as the zero vector in Axiom (V3). Finally, $(-1)\mathbf{f}$ acts as the opposite of $\mathbf{f}$. Axioms (V3) and (V4) are obviously true, and the other axioms follow from properties of the real numbers, so $X(\mathbb{D})$ is a vector space.

Other examples of vector spaces of functions are given by

- $C([a, b])$ the set of all continuous functions on $[a, b]$;
- $C^p([a, b])$ the set of all functions continuous up to their p-th derivative on $[a, b]$;
- $C^\infty(\mathbb{R})$ the set of all infinitely continuous on functions on $\mathbb{R}$.

5. Special attention is paid to the vector space $\mathbb{P}_n, n \geq 0$ of all polynomials of order n of the form

$$\mathbf{p}_n(t) = \sum_{k=0}^{n} a_k t^k, \tag{2.7}$$

where the coefficients $a_0, \ldots, a_n$ and the variable t are real numbers. The degree n of $\mathbf{p}$ is the highest power of t in (2.7), whose coefficient is not zero. The degree of a constant polynomial $\mathbf{p}(t) = a_0 \neq 0$ is zero. The zero polynomial has all the coefficients equal to zero and it is included in $\mathbb{P}_n$ even though its degree, for technical reasons, is not defined. The basic operations of addition and multiplication by scalar are defined pointwise as for functions.

Definition 2.3 (Vector subspace)**.** Let V be a vector space over a field $\mathbb{K}$. A non-empty subset $W \subseteq V$ is called a vector subspace if and only if is a vector space over $\mathbb{K}$ itself. In particular, only three of the vector space axioms need to be checked; the rest are automatically satisfied

(W1) $\forall \mathbf{w}_1, \mathbf{w}_2 \in W \implies \mathbf{w}_1 + \mathbf{w}_2 \in W$;

(W2) $\forall a \in \mathbb{K}, \forall \mathbf{w} \in W \implies a\mathbf{w} \in W$;

(W3) $\mathbf{0}_V \in W$.

Example 2.3 (Examples of vector subspaces)**.** Let us recall some remarkable examples of vector subspaces.

1. The set consisting of only the zero vector in a vector space V is a subspace of V, called the **zero subspace** and written as $\{\mathbf{0}_V\}$.

2. $C^p([a, b]) \subset C([a, b])$.

3. $\mathbb{P}_n \subset C^\infty(\mathbb{R})$.

4. The vector space $\mathbb{R}^2$ is **not** a subspace of $\mathbb{R}^3$ because $\mathbb{R}^2$ is not even a subset of $\mathbb{R}^3$ (vectors in $\mathbb{R}^3$ all have three components, whereas vectors in $\mathbb{R}^2$ have only two). We can define a subset of $\mathbb{R}^3$ which "looks" and "acts" like $\mathbb{R}^2$
$$H := \left\{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \in \mathbb{R}^3 : \ a, b \in \mathbb{R} \right\}. \tag{2.8}$$
It can be verified that H is a vector subspace of $\mathbb{R}^3$. For an exhaustive description of all subspaces of $\mathbb{R}^3$ see Exercise 2.2.

5. Let V be a vector space and let $U \subseteq V$, and $W \subseteq V$ be subspaces. We denote by $U \cap W$ the **intersection** of U and W, i.e. the set of elements which lie both in U and W
$$U \cap W := \{\mathbf{v} \in V : \mathbf{v} \in U \text{ and } \mathbf{v} \in W\}. \tag{2.9}$$
It can be verified that $U \cap W$ is a subspace. For instance, if U, W are two planes in $\mathbb{R}^3$ passing through the origin, then in general, their intersection will be a straight line passing through the origin.

6. Let V be a vector space and let $U \subseteq V$, and $W \subseteq V$ be subspaces. Let $U + W$ denote the set of all elements such that

$$U + W := \{\mathbf{v} \in V : \mathbf{v} = \mathbf{u} + \mathbf{w}, \mathbf{u} \in U, \mathbf{w} \in W\}. \tag{2.10}$$

It can be verified that $U + W$ is a subspace of V, said **to be generated** by U and W, and called the **sum** of U and W. If the expression as sum of an element of U and an element of W for any $\mathbf{v}$ in (2.10) is unique, than (2.10) is called **direct sum** and it is denoted by $U \oplus W$.

2.2 System of generators, linear dependence and independence of vectors.

Definition 2.4 (Span)**.** Let V be an arbitrary vector space and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ a set of n vectors in V. We call the subspace **generated** (or **spanned**) by $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ the set of all finite linear combinations of $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ defined by

$$\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} := \{a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \mid a_1, \ldots, a_n \in \mathbb{K}\}. \tag{2.11}$$

It can be easily verified that (2.11) defines a vector subspace of V.

Definition 2.5 (System of generators)**.** We say that the system $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ generates V if $\forall\, \mathbf{v} \in V$, there exist coefficients $a_1, \ldots, a_n \in \mathbb{K}$ such that

$$v = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n. \tag{2.12}$$

In other words, $V = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is called a **system of generators** for V.

Definition 2.6 (Linearly dependence and independence)**.** Let V be an arbitrary vector space and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ a set of n vectors in V. We say that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is **linearly independent** if the relation

$$V \ni \mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}_V \tag{2.13}$$

implies

$$a_1 = \cdots = a_n = 0_{\mathbb{K}}. \tag{2.14}$$

Otherwise, if there exist scalars $a_1, \ldots, a_n \in \mathbb{K}$, not all equal to 0, such that (2.13) holds true, we say that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is **linearly dependent**.

Example 2.4. Trivially, any set containing the vector $\mathbf{0}_V$ is linearly dependent.

Example 2.5. A set containing a single vector $\mathbf{v}$ is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}_V$.

Example 2.6. Trivially, if a set of vectors is linearly dependent, then one of them can be written as a linear combination of the others.

2.3 Bases and dimension of a vector space.

Definition 2.7 (Basis)**.** We call a **basis** of V any set of linearly independent vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ that generates V.

In particular, since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a system of generators for V, it follows that any $\mathbf{v} \in V$ can be written as

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n, \quad a_1, \ldots, a_n \in \mathbb{K}. \tag{2.15}$$

Then, the scalars $a_1, \ldots, a_n$ are called the **coefficients** (or **components** or **coordinates**) of $\mathbf{v}$ with respect to the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and decomposition (2.15) is called the **decomposition** of $\mathbf{v}$ w.r.t. the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Moreover, it can be proved that from $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ being also linearly independent, then the coefficients in decomposition (2.15) are uniquely defined, that is, if there exist coefficients $b_1, \ldots, b_n \in \mathbb{K}$ such that

$$\mathbf{v} = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n, \tag{2.16}$$

then necessarily

$$a_1 = b_1, \ldots, a_n = b_n. \tag{2.17}$$

This result is known in many textbooks as the "Unique Representation Theorem".

Example 2.7. Trivially, the vector $\mathbf{0}_V$ is never part of a basis.

The next theorem gives us useful information about the number of elements of any basis of a vector space. The main result is that any bases of a vector space have the same number of elements. Moreover, it gives us a useful criterion to determine when a set of elements of a vector space is a basis.

Theorem 2.1. Let V be a vector space over the field $\mathbb{K}$. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of V over $\mathbb{K}$. Then, the following statements hold true

1. any set of linearly independent vectors in V has at most n elements;
2. any other basis has exactly n elements;
3. any set of n linearly independent vectors in V must generate V, and hence, form a basis.

Remark 2.6. Theorem 2.1 suggests that there exists a number that is uniquely associated to any vector space V, which is the number of elements of any basis of V. This number is an intrinsic property of the space V that does not depend on the particular choice of the basis and that will define its dimension.

Definition 2.8 (Dimension of a vector space)**.** Let V be a vector space having a basis consisting of n elements. We call n the **dimension** of V and we write

$$\dim(\mathrm{V}) = n. \tag{2.18}$$

If V consists of $\mathbf{0}_V$ alone, then V does not have a basis, and we shall say that V has dimension 0. A vector space which has a basis consisting of a finite number of elements, or the zero vector space, is called **finite dimensional**. Otherwise, if for any n, there exist n linearly independent vectors of V, then V is called infinite dimensional and we write

$$\dim(\mathrm{V}) = \infty. \tag{2.19}$$

Theorem 2.2 (Dimension of a subspace)**.** Let V be a vector space having a basis of n elements. Let $W \subseteq V$ be a vector subspace. Then

$$\dim(\mathrm{W}) \leq \dim(\mathrm{V}). \tag{2.20}$$

Example 2.8 (Examples of dimensions of vector spaces)**.** Let us recall the dimensions of some remarkable vector spaces.

1. Let $\mathbb{K}$ be a field, which is a vector space over itself. Then its dimension is 1. Indeed, the element $1_{\mathbb{K}}$ forms a basis of $\mathbb{K}$ over $\mathbb{K}$, since any element $a \in \mathbb{K}$ has a unique expression as $a = a \cdot 1_{\mathbb{K}}$.

2. $\forall$ integer p, $\dim(C^p([a,b])) = \infty$;

3. $\dim(\mathbb{R}^n) = \dim(\mathbb{C}^n) = n$.

Remark 2.7 (Standard basis for $\mathbb{R}^n$)**.** The most commonly used basis for $\mathbb{R}^n$ is the **canonical** basis, or the **standard** basis, which is given by the set

$$\{\mathbf{e}_1 = (1,0,\ldots,0)^T, \mathbf{e}_2 = (0,1,\ldots,0)^T, \ldots, \mathbf{e}_n = (0,0,\ldots,1)^T\}. \tag{2.21}$$

In a compact form, it can defined in terms of the **Kronecker delta**

$$(\mathbf{e}_i)_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.22}$$

where $(\mathbf{e}_i)_j$ denotes the j-th component of $\mathbf{e}_i$, $\forall\, i, j = 1, \ldots, n$.

Remark 2.8. Generally, vectors are always intended as columns. For typographical reasons (it takes up less space), many texts write them as row vectors and use the symbol for the transpose to denote that they are column vectors.

2.4 Exercises

Exercise 2.1. Show that the vectors $(1,1)^T$ and $(-1,2)^T$ form a basis of $\mathbb{R}^2$.

Solution 2.1. Since $\dim(\mathbb{R}^2) = 2$, it is sufficient to prove that the two given vectors are linearly independent. Indeed, Theorem 2.1 implies that any other set of two linear independent vectors in $\mathbb{R}^2$ forms a basis. Consider a generic relation of type (2.13)

$$a_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + a_2 \begin{pmatrix} -1 \\ 2 \end{pmatrix} = \mathbf{0}_{\mathbb{R}^2} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{2.23}$$

for $a_1, a_2 \in \mathbb{R}$. Writing this equation in terms of components, we find that

$$\begin{cases} a_1 - a_2 = 0 \\ a_1 + 2a_2 = 0 \end{cases} \tag{2.24}$$

This is a system of two equations which we solve for a_1 and a_2. Subtracting the second from the first, we get $-3a_2 = 0$, whence $a_2 = 0$. Substituting in either equation, we find $a_1 = 0$. Hence a_1, a_2 are both 0, hence our vectors are linearly independent.

Exercise 2.2. Check if $\mathbf{v}_1 = (1,2)^T, \mathbf{v}_2 = (5,7)^T$ and $\mathbf{v}_3 = (10,5)^T$ are linearly independent or not.

Solution 2.2. Since $\dim(\mathbb{R}^2) = 2$, Theorem 2.1 implies that any set of linearly independent vectors in $\mathbb{R}^2$ has at most two elements, that is we can find at most two linearly independent vectors in $\mathbb{R}^2$. Here we are given three vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, which can not be linearly independent in $\mathbb{R}^2$.

Exercise 2.3. Find all possible vector subspaces of $\mathbb{R}^3$ with respective dimensions.

Solution 2.3. We first observe that by Definition 2.3 any subspace $W \subseteq \mathbb{R}^3$ must contain the vector $\mathbf{0}_{\mathbb{R}^3}$, that is the origin, and it has to satisfy the linearity conditions (W1) and (W2). In addition, Theorem 2.2 implies that there the dimension of any subspace has to be smaller or equal to three.

- *0-dimensional subspaces.* Only the zero subspace $\{\mathbf{0}_V\}$.
- *1-dimensional subspaces.* Any subspace spanned by a single nonzero vector

$$\text{span}\{\mathbf{v}\}, \mathbf{v} \neq \mathbf{0}_V. \tag{2.25}$$

  Such subspaces are lines through the origin.

- *2-dimensional subspaces.* Any subspace spanned by two linearly independent vectors $\mathbf{v}_1$ and $\mathbf{v}_2$

$$\text{span}\{\mathbf{v}_1, \mathbf{v}_2\}. \tag{2.26}$$

  Such subspaces are planes through the origin.

- *3-dimensional subspaces.* Only $\mathbb{R}^3$ itself. In particular, any three linearly independent vectors in $\mathbb{R}^3$ span all of $\mathbb{R}^3$ (cf. Theorem 2.1).

Remark 2.9. It is helpful to keep the geometric pictures of subspaces in Exercise 2.3 in mind, even for an abstract vector space.

Chapter 3

Matrices

3.1 The space of matrices

Definition 3.1 (Matrices)**.** Let $\mathbb{K}$ be a given field and let m, n be two positive integers. A **matrix** A over $\mathbb{K}$ having m **rows** and n **columns**, or an $m \times n$ (m-by-n) matrix is an array of mn scalars $a_{ij} \in \mathbb{K}$, $i = 1, \ldots, m$ and $j = 1, \ldots, n$, represented as

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \tag{3.1}$$

The index i denotes the rows and the index j denotes the columns. For instance, the $i-$th row is $\begin{bmatrix} a_{i1}, a_{i1}, \ldots, a_{in} \end{bmatrix}$ and the $j-$th column is

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \ldots \\ a_{mj} \end{bmatrix} \tag{3.2}$$

We abbreviate the notation (3.1) by $A = (a_{ij})$, $1 = 1, \ldots, m$ and $j = 1, \ldots n$ and we write $A \in \mathbb{K}^{m \times n}$.

Definition 3.2 (Basic terminology)**.** The integers m and n are called the **dimensions** of the matrix A and the couple (m, n) defines its **size** (or **shape**). The scalars a_{ij} are called its **entries** or **components**. The **diagonal entries** are $a_{11}, a_{22}, a_{33}, \ldots$ and they form the **main diagonal** of A. If $m = n$, then the matrix A is called a **square matrix**. A **diagonal matrix** is a square matrix whose nondiagonal entries are zero, i.e. $a_{ij} = 0$ for $i \neq j$. An example is the $n \times n$ **identity matrix** (or the **identity matrix of order** n)

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}. \tag{3.3}$$

An $m \times n$ matrix whose entries are all zero is a **zero matrix** and it is written as O.

$$O = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \in \mathbb{K}^{m\times n}. \tag{3.4}$$

The size of a zero matrix is usually clear from the context.

We say that two matrices $A = (a_{ij})$ and $B = (b_{ij})$ are **equal** if they have the same size (i.e., the same number of rows and the same number of columns) and their corresponding entries are equal, i.e. $a_{ij} = b_{ij}$, $\forall i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$.

Example 3.1 (Examples of matrices)**.** Let us consider some concrete examples.

1. The following is a 3×2 matrix with real entries:

$$A = \begin{bmatrix} 5 & 2 \\ 7 & -7 \\ 0 & 8 \end{bmatrix} \in \mathbb{R}^{3\times 2}. \tag{3.5}$$

   A is an array of $3 \cdot 2 = 6$ real numbers. The rows are $[5, 2]$, $[7, -7]$ and $[0, 8]$ and the columns are

$$\begin{bmatrix} 5 \\ 7 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -7 \\ 8 \end{bmatrix}. \tag{3.6}$$

2. An example of a matrix over the field of complex numbers is:

$$B = \begin{bmatrix} i & 0 & 7 & 2i \\ 2+1 & 4 & 3 & 2-i \end{bmatrix} \in \mathbb{C}^{2\times 4}. \tag{3.7}$$

3. A row vector $[x_1, x_2, \dots, x_n] \in \mathbb{R}^{1\times n}$ is a particular $1 \times n$ matrix, while a column vector $[y_1, y_2, \dots, y_m]^T \in \mathbb{R}^{m\times 1}$ is a particular $m \times 1$ matrix. A single scalar $[a]$ can be viewed as a 1×1 matrix.

3.2 Matrix operations

We shall now define the operations of matrix addition and matrix multiplication by scalars. The basic arithmetic for matrices is a natural extension of the basic arithmetic for scalars.

Definition 3.3 (Matrix addition)**.** Given $A = (a_{ij})$ and $B = (b_{ij})$ two $m \times n$ matrices over $\mathbb{K}$, we define the **sum** of A and B to be the $m \times n$ matrix $A + B$ obtained by adding corresponding entries. That is,

$$(A + B)_{ij} = a_{ij} + b_{ij} \, \forall i = 1, \ldots, m, \text{ and } \forall j = 1, \ldots, n. \tag{3.8}$$

For example,

$$\begin{bmatrix} 2 & 1 & 3 \\ 0 & -7 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 3 \\ 1 & -1 & -2 \end{bmatrix} = \begin{bmatrix} 2+1 & 1+0 & 3+3 \\ 0+1 & -7-1 & 0-2 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 6 \\ 1 & -8 & -2 \end{bmatrix}. \tag{3.9}$$

Note that addition is only defined for matrices having the same shape.

Remark 3.1. The symbol "+" is used to denote addition between scalars in some places and addition between matrices at other places. Although these are two distinct algebraic operations, no ambiguities will arise from the context.

Remark 3.2. The matrix $(-A)$, called the **additive inverse** of A, is defined to be the matrix obtained by negating each entry of A. That is, if $A = (a_{ij})$, then $-A = (-a_{ij})$. This allows matrix subtraction to be defined in the natural way. For two matrices of the same shape, the difference $A - B$ is defined to be the matrix $(A - B)_{ij} = a_{ij} - b_{ij}$, $\forall i = 1, \ldots, m$, and $\forall j = 1, \ldots, n$.

Since matrix addition is defined in terms of scalar addition, the familiar algebraic properties of scalar addition are inherited by matrix addition. Indeed, for $m \times n$ matrices A, B, and C, the following properties hold.

- Associative property: $(A + B) + C = A + (B + C)$;
- Commutative property: $A + B = B + A$;
- The $m \times n$ zero matrix O has the property that $A + O = A$.
- The $m \times n$ matrix $(-A)$ has the property that $A + (-A) = O$.

Definition 3.4 (Matrix multiplication by scalars)**.** Given $A = (a_{ij})$ an $m \times n$ matrix over $\mathbb{K}$ and a scalar $c \in \mathbb{K}$, we define the **multiplication** of A by c to be the $m \times n$ matrix cA obtained by multiplying each entry of A by c. That is,

$$(cA)_{ij} = ca_{ij} \, \forall i = 1, \ldots, m, \text{ and } \forall j = 1, \ldots, n. \tag{3.10}$$

For example,

$$5 \begin{bmatrix} -i & 3 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} -5i & 15 \\ 0 & 15 \end{bmatrix}. \tag{3.11}$$

The rules for combining addition and multiplication by scalars between matrices are listed below. Indeed, for $m \times n$ matrices A and B and for scalars c and d, the following properties hold.

- Associative property: $(cd)A = c(dA)$;
- Distributive property: $c(A + B) = cA + cB$ (multiplication by scalar is distributed over matrix addition);
- Distributive property: $(c + d)A = cA + dA$ (multiplication by scalar is distributed over scalar addition);
- $1_{\mathbb{K}} A = A$.

Remark 3.3 (Vector space over $\mathbb{K}$)**.** Let $\mathcal{M}at(m, n; \mathbb{K})$ denote the set of all $m \times n$ matrices over a field $\mathbb{K}$. The properties of matrix addition and multiplication by scalars show that it forms a vector space over $\mathbb{K}$. We shall equivalently denote it by $\mathbb{K}^{m\times n}$ or by $\mathcal{M}_{m\times n}(\mathbb{K})$. Its dimension is mn.

Remark 3.4 (Canonical basis)**.** For $\mathcal{M}at(m, n; \mathbb{R})$ the *canonical basis* is given by all the $m \times n$ matrices $E_{ij} = (e_{ij})\, i, = 1, \ldots m$ and $j = 1, \ldots, n$ whose entries are all equal to zero, a part from the entry e_{ij}, which is equal to one.

Example 3.2. The space of all 2×2 real matrices has dimension four. Its canonical basis has four elements

$$E_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \text{ and } E_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Every 2×2 matrix can be decomposed with respect to this basis. For example,

$$A = \begin{bmatrix} 3 & 4 \\ 7 & 0 \end{bmatrix} = 3 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 4 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 7 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 0 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \tag{3.12}$$

3.3 Matrix multiplication

The following sections are devoted to the introduction and discussion of other operations between matrices that don't derive from basic operations between scalars. Such operations are matrix multiplication, transposition and inversion.

Definition 3.5 (Matrix multiplication)**.** We first introduce a new notion.

- Two matrices A and B are said to be **compatible** (or **conformable**) for multiplication AB whenever the number of columns of A is the same as the number of rows of B, i.e. A is $m \times n$ and B is $n \times p$.

- For compatible matrices $A = (a_{ij}) \in \mathbb{K}^{m \times n}$ and $B = (b_{ij}) \in \mathbb{K}^{n \times p}$, the **matrix product** AB is defined to be the $m \times p$ matrix whose (i, j)-entry is the inner product of the i^{th} row of A with the j^{th} column in B. That is, for $i = 1, \dots, m,$ and $k = 1, \dots, p$

$$(AB)_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \dots + a_{in}b_{nk} = \sum_{j=1}^{n} a_{ij}b_{jk}. \tag{3.13}$$

- In case A and B fail to be conformable, then no product AB is defined.

Example 3.3. If

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & -1 \\ 4 & 0 \end{bmatrix}, \tag{3.14}$$

we can compute AB and it is going to be a 2×2 matrix

$$AB = \begin{bmatrix} 2 \cdot 1 + 1 + 4 & 2 \cdot (-1) + 1 \cdot 0 \\ 0 \cdot 1 + 3 \cdot 3 & 0 \cdot (-1) + 3 \cdot 0 \end{bmatrix} = \begin{bmatrix} 7 & -2 \\ 9 & 0 \end{bmatrix}. \tag{3.15}$$

Example 3.4. Consider the matrices

$$A = \begin{bmatrix} 7 & 2 \\ 1 & 1 \\ 0 & 5 \\ -3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 3 & 4 & 1 \\ 0 & 0 & 4 \end{bmatrix}. \tag{3.16}$$

Matrices compatible for multiplication are A and C, and C and B. All the other combinations –i.e., A and B, B and A, and B and C– don't lead to compatibility for multiplication. We can compute

$$AC = \begin{bmatrix} 7 \cdot 3 + 2 \cdot 0 & 7 \cdot 4 + 2 \cdot 0 & 7 \cdot 1 + 2 \cdot 4 \\ 1 \cdot 3 + 1 \cdot 0 & 1 \cdot 4 + 1 \cdot 0 & 1 \cdot 1 + 1 \cdot 4 \\ 0 \cdot 3 + 5 \cdot 0 & 0 \cdot 4 + 5 \cdot 0 & 0 \cdot 1 + 5 \cdot 4 \\ -3 \cdot 3 + 4 \cdot 0 & -3 \cdot 4 + 4 \cdot 0 & -3 \cdot 1 + 4 \cdot 4 \end{bmatrix} = \begin{bmatrix} 21 & 28 & 15 \\ 3 & 4 & 5 \\ 0 & 0 & 20 \\ -9 & -12 & 13 \end{bmatrix} \tag{3.17}$$

and

$$CB = \begin{bmatrix} 3 \cdot 1 + 4 \cdot 0 + 1 \cdot 0 & 3 \cdot 0 + 4 \cdot 1 + 1 \cdot 0 & 3 \cdot 0 + 4 \cdot 0 + 1 \cdot 1 \\ 0 \cdot 1 + 0 \cdot 0 + 4 \cdot 0 & 0 \cdot 0 + 0 \cdot 1 + 4 \cdot 0 & 0 \cdot 0 + 0 \cdot 0 + 4 \cdot 1 \end{bmatrix} = \begin{bmatrix} 3 & 4 & 1 \\ 0 & 0 & 4 \end{bmatrix}. \tag{3.18}$$

Remark 3.5. Example 3.4 shows that even if the product AB exists, the product BA might not be defined. If particular, if A is $m \times n$, then B needs to be $n \times m$ in order for both products AB and BA to exist.

Remark 3.6 (Matrix multiplication is not commutative.)**.** Moreover, even when both products AB and BA exist, they need not be equal. In general, matrix multiplication is a noncommutative operation, i.e.

$$AB \neq BA. \tag{3.19}$$

For example, if

$$A = \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, \tag{3.20}$$

then

$$AB = \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix} \quad \text{but} \quad BA = \begin{bmatrix} 3 & -2 \\ 6 & -4 \end{bmatrix} \neq AB. \tag{3.21}$$

Example 3.5. For scalars, $ab = 0$ implies $a = 0$ or $b = 0$. The following example shows that the analogous statement for matrices does not hold—the matrices in this example show that it is possible for $AB = O$ with $A \neq O$ and $B \neq O$. Let

$$A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \tag{3.22}$$

then

$$AB = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{3.23}$$

in spite of $A \neq O$ and $B \neq O$.

Example 3.6. Recall the **cancellation law** for scalars, which states that $ab = ac$ and $a \neq 0$ implies $b = c$. A similar row does not hold for matrices. In fact, if

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \tag{3.24}$$

then

$$AB = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} = AC \quad \text{but} \quad B \neq C \tag{3.25}$$

in spite of the fact that $A \neq O$.

Although that there are some differences between scalar and matrix algebra –most notable is the fact that matrix multiplication is not commutative, and there is no cancellation law– fortunately, other important properties hold for matrix multiplication. For compatible matrices, the following properties hold true

- Left-hand distributive property: $A(B+C) = AB + BC$;
- Right-hand distributive property: $(D+E)F = DF + EF$;
- Associative property: $A(BC) = (AB)C$.

Remark 3.7 (Identity matrix)**.** We have already defined in Equation (3.3) the identity matrix of order n with ones on the main diagonal and zeros elsewhere. It plays the "role" of the identity element for matrix multiplication because it has the property that it reproduces whatever it is multiplied by. Indeed, for every $m \times n$ matrix A,

$$AI_n = A \quad \text{and} \quad I_m A = A. \tag{3.26}$$

Definition 3.6 (Powers of matrices)**.** Given a square $n \times n$ matrix A, we define the 0^{th} power of A to be the identity matrix of order n, that is

$$A^0 = I_n. \tag{3.27}$$

For a nonnegative integer s, we define

$$A^s = \underbrace{AA \cdots A}_{\text{s times}}. \tag{3.28}$$

The associative law guarantees that it makes no difference how matrices are grouped for powering, and due to the lack of compatilibity, powers of nonsquare matrices are never defined. Also, the usual laws of exponents hold. For nonnegative integers r and s,

$$A^r A^s = A^{r+s} \quad \text{and} \quad (A^r)^s = A^{rs}. \tag{3.29}$$

3.4 Matrix inversion

Definition 3.7 (Matrix inverse)**.** Given a square matrix A of order n, we call the **inverse** of A and denote it by A^{-1}, the square matrix of order n that satisfies the conditions

$$AA^{-1} = I_n \quad \text{and} \quad A^{-1}A = I_n. \tag{3.30}$$

Notice that condition (3.30) rules out inverses of nonsquare matrices.

Remark 3.8. Not all square matrices are invertible, i.e., given a square matrix A of order n, we can not always find a square matrix A^{-1} of order n satisfying conditions (3.30). A trivial example of is the zero matrix or order n, but there are also many nonzero matrices that are not invertible.

Remark 3.9. Although not all matrices are invertible, *when an inverse exists, it is unique.*

Definition 3.8. An invertible matrix is said to be **nonsingular**, and a square matrix with no inverse is called a **singular** matrix.

For nonsingular matrices A and B, the following properties hold.

– $(A^{-1})^{-1} = A$;

– The product AB is also nonsingular;

– $(AB)^{-1} = B^{-1}A^{-1}$ (the reverse order property for inversion).

More generically, if $A_1, A_2, \ldots, A_k$ are each $n \times n$ nonsingular matrices, then the product $A_1A_2 \ldots A_k$ is also nonsingular, and its inverse is given by the reverse order property

$$(A_1A_2 \ldots A_k)^{-1} = A_k^{-1} \ldots A_2^{-1}A_1^{-1}. \tag{3.31}$$

Example 3.7 (Inverse of a 2×2 matrix)**.** Given a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \text{where} \quad \delta := ad - bc \neq 0, \tag{3.32}$$

then

$$A^{-1} = \frac{1}{\delta} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{3.33}$$

because it can be easily verified that A^{-1} satisfies condition (3.30). We shall see in Section 3.6 that the number δ defines the determinant of A. More generically, a matrix will be invertible if and only if its determinant is different from zero.

3.5 Matrix transposition

Another matrix operation that is not derived from scalar arithmetic is matrix transposition.

Definition 3.9 (Matrix transpose)**.** Given an $m \times n$ matrix A over $\mathbb{R}$, the **transpose** of A is defined to be the $n \times m$ matrix A^T obtained by interchanging rows and columns in A. More precisely, if $A = (a_{ij})$, then

$$(A^T)_{ij} = a_{ji}, \ \forall i = 1, \ldots, m, \ \text{ and } j = 1, \ldots, n. \tag{3.34}$$

For example,

$$\begin{bmatrix} 7 & 2 \\ 1 & 0 \\ 0 & 5 \\ -3 & 4 \end{bmatrix}^T = \begin{bmatrix} 7 & 1 & 0 & -3 \\ 2 & 0 & 5 & 4 \end{bmatrix}. \tag{3.35}$$

Trivially, for all matrices, $(A^T)^T = A$.

Definition 3.10 (Matrix adjoint)**.** Given an $m \times n$ matrix A over $\mathbb{C}$, the **conjugate transpose** (or **adjoint**) of A is defined to be the $n \times m$ matrix A^H (sometimes also denoted by $A^\star$) obtained by interchanging rows and columns in A accompanied by the transpose operation of the complex entries of A. Recall that the complex conjugate of the complex number $z = a + ib$ is defined to be the complex number $\bar{z} = a - ib$. In other words, for $A = (a_{ij})$, then

$$(A^H)_{ij} = \bar{a}_{ji}, \ \forall i = 1, \ldots, m, \ \text{ and } j = 1, \ldots, n. \tag{3.36}$$

For example,

$$\begin{bmatrix} 2-3i & 2 \\ -i & 1 \\ 0 & 5 \\ -3 & -8-4i \end{bmatrix}^H = \begin{bmatrix} 2+3i & i & 0 & -3 \\ 2 & 1 & 5 & -8+4i \end{bmatrix}. \tag{3.37}$$

The transpose (and conjugate transpose) operation is easily combined with matrix addition and scalar multiplication. Indeed, given A and B two matrices of the same shape over $\mathbb{K}$ ($K = \mathbb{R}$ or $K = \mathbb{C}$) and c a scalar, then the following properties hold true.

- $(A + B)^T = A^T + B^T \quad \text{and} \quad (A + B)^H = A^H + B^H$;
- $c(A)^T = A^T + B^T \quad \text{and} \quad (A + B)^H = A^H + B^H$.

Definition 3.11 (Symmetries)**.** Given a square matrix $A = (a_{ij})$ over $\mathbb{R}$.

- A is said to be a **symmetric** matrix whenever $A = A^T$, i.e., whenever $a_{ij} = a_{ji}$.

- A is said to be a **skew-symmetric** matrix whenever $A = -A^T$, i.e., whenever $a_{ij} = -a_{ji}$.

Given a square matrix $A = (a_{ij})$ over $\mathbb{C}$.

- A is said to be a **hermitian** matrix whenever $A = A^H$, i.e., whenever $a_{ij} = \bar{a}_{ji}$.
- A is said to be a **skew-hermitian** matrix whenever $A = -A^T$, i.e., whenever $a_{ij} = -\bar{a}_{ji}$.

3.6 Determinants

We start by a gentle introduction on the topic by first carrying out separately the computation of determinants of order 2 and 3. After that, we introduce the generalised theory for determinants of order n.

Definition 3.12 (Determinants of order 2)**.** Given a 2×2 matrix over a field $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$ for us)

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \tag{3.38}$$

the determinant of A is defined to be the number

$$\det(A) := a_{11}a_{22} - a_{12}a_{21}. \tag{3.39}$$

For example, if

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 3 \end{bmatrix}, \tag{3.40}$$

then

$$\det(A) = 1 \cdot 3 - 0 \cdot (-1) = 3. \tag{3.41}$$

Remark 3.10. We shall equivalently use the notation $|A|$ or $\begin{vmatrix} x & y \\ z & v \end{vmatrix}$ for $\det(A)$.

Exercise 3.1. Show that if the columns (or rows) of A are interchanged, the determinant (3.39) changes by a sign.

Solution 3.1. By direct application of (3.39), we find that if we interchange the columns

$$\begin{vmatrix} a_{12} & a_{11} \\ a_{22} & a_{21} \end{vmatrix} = a_{12}a_{21} - a_{11}a_{22} = -(a_{11}a_{22} - a_{21}a_{12}) = -\det(A). \tag{3.42}$$

Equivalently, when rows are interchanged

$$\begin{vmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{vmatrix} = a_{21}a_{12} - a_{11}a_{22} = -(a_{11}a_{22} - a_{12}a_{21}) = -\det(A). \tag{3.43}$$

Exercise 3.2. Show that the columns of A are linearly dependent if and only if the determinant (3.39) is zero (equivalently, the columns of A are linearly independent if and only if $\det(A) \neq 0$).

Solution 3.2. By definition, the columns of A are linearly dependent if there exist scalars $(c, d) \neq (0, 0)$, such that

$$c\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + d\begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix} = \begin{pmatrix} ca_{11} + da_{12} \\ ca_{21} + da_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{3.44}$$

Writing this equation in terms of components and assuming $c \neq 0$, we find a system of two equations which we solve for a_{11} and a_{21}.

$$\begin{cases} a_{11} = -(d/c)a_{12} \\ a_{21} = -(d/c)a_{22} \end{cases} \tag{3.45}$$

By substituting the entries of A, we compute

$$\det(A) = \begin{vmatrix} -(d/c)a_{12} & a_{12} \\ -(d/c)a_{22} & a_{22} \end{vmatrix} = -(d/c)a_{12}a_{22} - a_{12}(-(d/c)a_{22}) = 0. \tag{3.46}$$

Exercise 3.3. Show that if two columns (or rows) of A are equal, the determinant (3.39) is zero.

Solution 3.3. By direct application of (3.39), we find

$$\begin{vmatrix} a & a \\ b & b \end{vmatrix} = ab - ab = 0 \quad \text{and} \quad \begin{vmatrix} a & b \\ a & b \end{vmatrix} = ab - ab = 0.$$

Exercise 3.4. Show that the determinant of the identity matrix of order 2 is 1.

Solution 3.4. Trivially,

$$\det(I_2) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 - 1 = 0.$$

Remark 3.11. Exercises 3.1, 3.2, 3.3 and 3.4 can be generalised to any dimension $n \geq 1$ and hold true also for higher dimensions. They represent some of the main properties of determinants.

Definition 3.13 (Determinants of order 3)**.** Given a 3×3 matrix over $\mathbb{K}$

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \tag{3.47}$$

its determinant can be computed using the *expansion by the first row* rule

$$\begin{aligned}\det(A) &= a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13}) \\ &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ \alpha_{32} & a_{33}\end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ \alpha_{31} & a_{33}\end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ \alpha_{31} & a_{32}\end{vmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.\end{aligned} \tag{3.48}$$

In (3.48) A_{ij} represents the matrix obtained by eliminating from A the i^{th} row and the j^{th} column.

Remark 3.12. We can expand by any row and column following the pattern of the signs

$$\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix} \tag{3.49}$$

and the result (3.48) won't change.

Remark 3.13. In general, when an entry is equal to zero it is convenient to expand according to the corresponding row or a column. For example, by expanding by the second column in the following example, we find

$$\begin{aligned}\begin{vmatrix} 3 & 0 & 1 \\ 1 & 2 & 5 \\ -1 & 4 & 2\end{vmatrix} &= -0\begin{vmatrix} 1 & 5 \\ -1 & 2\end{vmatrix} + 2\begin{vmatrix} 3 & 1 \\ -1 & 2\end{vmatrix} - 4\begin{vmatrix} 3 & 1 \\ 1 & 5\end{vmatrix} \\ &= 14 - 4\cdot 14 \\ &= -42.\end{aligned} \tag{3.50}$$

The expansion by a row or a column rules are particular cases of the generic **Laplace rule for the expansion of the determinant**, which allows to reduce the computation of an $n \times n$ determinant to that of n $(n-1)\times(n-1)$ determinants. Before introducing that, let us give the general definition of determinants of order n.

Recall that a **permutation** $\pi = (\pi_1, \ldots, \pi_n)$ of the numbers $(1, 2, \ldots, n)$ is simply any rearrangement. In general, there exist has $n! = n(n-1)(n-2)\ldots 1$ different permutations of the sequence $(1, 2, \ldots, n)$.

The **sign of a permutation** π is defined to be the number

$$\operatorname{sgn}(\pi) = \begin{cases} +1 \text{ if an even number of exchanges is needed to restore } \pi \text{ to natural order} \\ -1 \text{ if an odd number of exchanges is needed to restore } \pi \text{ to natural order}\end{cases}$$

For example, if $p = (1, 4, 3, 2)$ is a permutation of $(1, 2, 3, 4)$, then $\text{sgn}(p) = -1$, and if $p = (4, 3, 2, 1)$, then $\text{sgn}(p) = +1$. The sign of the natural order $p = (1, 2, 3, 4)$ is naturally $\text{sgn}(p) = +1$. The general definition of the determinant can now be given.

Definition 3.14 (Determinants of order n)**.** Let $A = (a_{ij})$ be a square matrix of order n over $\mathbb{K}$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \tag{3.51}$$

the **determinant** of A is defined to be the scalar

$$\det(A) = \sum_{\pi \in P} \text{sgn}(\pi) a_{1\pi_1} a_{2\pi_2} \dots a_{n\pi_n} \in \mathbb{K}, \tag{3.52}$$

where $P = \{\pi = (\pi_1, \dots, \pi_n)\}$ is the set of $n!$ permutations of $(1, \dots, n)$. Observe that each term $a_{1\pi_1} a_{2\pi_2} \dots a_{n\pi_n}$ in (3.52) contains exactly one entry from each row and each column of A. Note that the determinant is only defined for square matrices.

Exercise 3.5. Compute the determinant of a 2×2 matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ by applying the definition.

Solution 3.5. There are $2! = 2$ permutations of the sequence $(1, 2)$, and precisely $P = \{(1, 2), (2, 1)\}$, so $\det(A)$ contains the two terms

$$\text{sgn}(1, 2) a_{11} a_{22} \quad \text{and} \quad \text{sgn}(2, 1) a_{12} a_{21}. \tag{3.53}$$

In particular, $\text{sgn}(1, 2) = +1$ and $\text{sgn}(2, 1) = -1$. By substituting, we obtain the already familiar formula (3.39)

$$\det(A) = a_{11} a_{22} - a_{12} a_{21}. \tag{3.54}$$

Definition 3.15 (Laplace's expansion formula for the determinant)**.** Laplace's expansion rule for the computation of an $n \times n$ determinant according the i^{th} row is

$$\det(A) = \begin{cases} a_{11} \text{ if } n = 1 \\ \sum_{j=1}^{n} (-1)^{i+j} \det(A_{ij}) a_{ij} \text{ if } n \geq 2 \end{cases} \tag{3.55}$$

and according to j^{th} column

$$\det(A) = \begin{cases} a_{11} \text{ if } n = 1 \\ \sum_{i=1}^{n} (-1)^{i+j} \det(A_{ij}) a_{ij} \text{ if } n \geq 2 \end{cases} \tag{3.56}$$

The scalar $\det(A_{ij})$ is called the **complementary minor** associated with a_{ij} and the term $C_{ij} := (-1)^{i+j}\det(A_{ij})$ is called the **cofactor** of the entry a_{ij}.

Let us enumerate the main properties that are satisfied by determinants of any order:

- $\det(A^T) = \det(A)$, $\forall n$;
- $\det(I_n) = 1$, $\forall n$;
- If two columns or rows are interchanged, the determinant changes by a sign;
- If the matrix B is obtained by multiplying row i of A by the scalar c, then $\det(B) = c\det(A)$;
- If the matrix B is obtained by adding c times row i of A to row j of A, then $\det(B) = \det(A)$;
- An $n \times n$ matrix A is nonsingular if and only if $\det(A) \neq 0$ (or, equivalently, A $n \times n$ is singular if and only if $\det(A) = 0$).
- For any $n \times n$ matrix A and scalar c, $\det(cA) = c^n \det(A)$;
- For compatible matrices A and B, $det(AB) = \det(A)\det(B)$ (*Binet's formula*).

Chapter 4

Linear mappings

4.1 The space of linear mappings

Definition 4.1. Given two vector spaces V and W over $\mathbb{K}$, a mapping $L : V \to W$ is **linear** if it *preserves the operations of vector addition and multiplication by scalar*, that is

i) $L(\mathbf{v}_1 + \mathbf{v}_2) = L(\mathbf{v}_1) + L(\mathbf{v}_2)$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$;

ii) $L(c\mathbf{v}) = cL(\mathbf{v})$ for all $c \in \mathbb{K}$ and for all $\mathbf{v} \in V$;

or, equivalently, both conditions can be combined into one requirement

iii) $L(c\mathbf{v}_1 + d\mathbf{v}_2) = cL(\mathbf{v}_1) + dL(\mathbf{v}_2)$ for all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and for all $c, d \in \mathbb{K}$.

In particular, condition ii) implies that

$$L(\mathbf{0}_V) = \mathbf{0}_W, \tag{4.1}$$

since $L(\mathbf{0}_V) = L(0_{\mathbb{K}}\mathbf{v}) = 0_{\mathbb{K}}L(\mathbf{v}) = \mathbf{0}_W$ for all $\mathbf{v} \in V$.

Repeated application of condition iii) produces the generalization

$$L(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k) = c_1L(\mathbf{v}_1) + c_2L(\mathbf{v}_2) + \cdots + c_kL(\mathbf{v}_k) \tag{4.2}$$

for all $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_k \in V$ and for all $c_1, c_2, \ldots, c_k \in \mathbb{K}$.

Suppose that we are given $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ a basis for V and we want to know the value of the linear mapping for a generic vector $\mathbf{v} \in V$. The crucial consequence of linearity is that *if we know the value of the mapping for each vector in the basis, then we know its value for each vector in the entire space.* Indeed, since $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis, then each $\mathbf{v}$ in V can be uniquely written as

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n, \tag{4.3}$$

for unique coefficients $c_1, c_2, \ldots, c_n \in \mathbb{K}$. The generalized linearity condition (4.2) implies that

$$\begin{aligned} L(\mathbf{v}) &= L(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) \\ &= c_1 L(\mathbf{v}_1) + c_2 L(\mathbf{v}_2) + \cdots + c_n L(\mathbf{v}_n). \end{aligned} \tag{4.4}$$

The only freedom for the mapping is its value on the basis vectors. Once we settle the basis vectors, the mapping of every vector of the space is settled.

Remark 4.1. The addition symbol "+" in conditions i) and iii) in some places denotes addition between vectors in V "$+_V$" (e.g, in $L(\mathbf{v}_1 +_V \mathbf{v}_2)$) and addition between vectors in W "$+_W$" in others (e.g., in $L(\mathbf{v}_1) +_W L(\mathbf{v}_2)$). We shall write simply "+" and we shall assume the space is clear from the context.

Definition 4.2 (Basic terminology)**.** Linear mappings are often called linear **transformations** or linear **applications**. A mapping will also be called a map, for the sake of brevity. A linear map from V into itself is called a linear **operator** on V. Like for classic terminology for functions, the vector space V is called the **domain** of T and W is called the **codomain** of T. For $\mathbf{v} \in V$, the vector $T(\mathbf{v})$ is called the **image** of $\mathbf{v}$ under the action of T.

Definition 4.3 (The space of linear maps)**.** Let V and W be two vector spaces over $\mathbb{K}$. We denote the space of all linear maps by

$$\mathcal{L}(V, W) := \{L : V \to W \text{ linear mapping}\}. \tag{4.5}$$

We shall define the operations of addition multiplication by scalars in $\mathcal{L}(V, W)$ in such a way as to make it into a vector space.

Given $L, F \in \mathcal{L}(V, W)$, their **addition** $L + F$ is the map $L + F : V \to W$, such that

$$(L + F)(\mathbf{v}) = L(\mathbf{v}) + F(\mathbf{v}) \quad \text{for each } \mathbf{v} \in V. \tag{4.6}$$

It can be easily verified that the map $L + F$ satisfies the two conditions i) and ii) which define a linear map.

Furthermore, if $a \in \mathbb{K}$ and $L \in \mathcal{L}(V, W)$, we define the map $(aL) : V \to W$, such that

$$(aL)(\mathbf{v}) = aL(\mathbf{v}) \quad \text{for each } \mathbf{v} \in V. \tag{4.7}$$

Again, it can be easily verified that aL is a linear map.

Given $L \in \mathcal{L}(V, W)$, we can define the opposite map $-T$ to be the map $(-1)T$. Finally, we recall that we have already defined the zero map. We can easily verify that the properties (V1) through (V8) are satisfied and conclude that the set of linear maps between two vector spaces is itself a vector space. Its dimension is

$$\dim(\mathcal{L}(V, W)) = \dim(V) \cdot \dim(W). \tag{4.8}$$

4.2 Image and Kernel of linear mappings

Definition 4.4 (Kernel of a linear map)**.** Let V and W be vector spaces over $\mathbb{K}$, and let $F : V \to W$ be a linear map. We define the **kernel** of F to be the set of all elements in V, whose image in W is the zero vector

$$\mathrm{Ker}(F) := \{\mathbf{v} \in V : F(\mathbf{v}) = \mathbf{0}_W\} \subseteq V. \tag{4.9}$$

It is always a non-empty set, since $\mathbf{0}_V$ is always an element. It can be proved that the kernel of a linear map F is a vector subspace of V.

The kernel of a linear map is useful to determine when the map is injective.

Definition 4.5 (Injective map)**.** A linear mapping $F : V \to W$ is **injective** if different vectors in V are mapped into different vectors in W, that is, if $\mathbf{v}_1, \mathbf{v}_2$ are elements of V such that $F(\mathbf{v}) = F(\mathbf{w})$, then $\mathbf{v} = \mathbf{w}$.

Theorem 4.1. Let $F : V \to W$ be a linear map. The following two conditions are equivalent:

1. $\mathrm{Ker}(F) = \{\mathbf{0}_V\}$.
2. F is injective.

Theorem 4.2. Let $F : V \to W$ be an injective linear map. If $\mathbf{v}_l, \dots, \mathbf{v}_n$ are linearly independent elements of V, then $L(\mathbf{v}_l), \dots, L(\mathbf{v}_n)$ are also linearly independent elements of W.

Definition 4.6 (Image of a linear map)**.** Let $F : V \to W$ be a linear map. We define the **image** of F to be the set of all elements in W that are the image through F of at least one element in V

$$\mathrm{Im}(F) = \{\mathbf{w} \in W : \exists\, \mathbf{v} \in V : F(\mathbf{v}) = \mathbf{w}\} \subseteq W. \tag{4.10}$$

The image of F is a vector subspace of W. In particular, it is always non-empty since the zero vector in V is always mapped to the zero vector in W.

Definition 4.7 (Surjective map)**.** A linear $F : V \to W$ is **surjective** if the domain V is mapped through F to the entire codomain W, that is $\mathrm{Im}(V) = W$.

The next theorem relates the dimensions of the kernel and image of a linear map with the dimension of the space on which the map is defined.

Theorem 4.3 (Rank-Nullity Theorem)**.** Let V be a vector space. Let $F : V \to W$ be a linear map of V into another vector space W. Then

$$\dim(V) = \dim(\mathrm{Ker}(F)) + \dim(\mathrm{Im}(F)). \tag{4.11}$$

Definition 4.8 (**Bijective map**). A linear $F : V \to W$ is **bijective** if it is injective and surjective.

Theorem 4.4. Let V and W be two vector spaces for which $\dim(V) = \dim(W)$. Let $F : V \to W$ be a linear map. If F is injective, or if F is surjective, then F is bijective.

4.3 Linear mappings and matrices

Definition 4.9 (The linear map associated to a matrix). Given an $m \times n$ matrix A over $\mathbb{K}$, we can always associate with A a linear map

$$L_A : \mathbb{K}^n \to \mathbb{K}^m \tag{4.12}$$

by defining

$$L_A(\mathbf{v}) := A\mathbf{v}, \tag{4.13}$$

where the product in (4.13) represents the matrix product by seeing any column vector $\mathbf{v} \in \mathbb{K}^n$ as an $n \times 1$ matrix. Both A and v are compatible for multiplication and the result of the product will be an $m \times 1$ matrix, that is a column vector in $\mathbb{K}^m$. Linearity is an immediate consequence of properties of matrix multiplication described in Section 3.3. Indeed, we have that

$$L_A(\mathbf{v}_1 + \mathbf{v}_2) = A(\mathbf{v}_1 + \mathbf{v}_2) = A\mathbf{v}_1 + A\mathbf{v}_2 = L_A(\mathbf{v}_1) + L_A(\mathbf{v}_2) \tag{4.14}$$

and

$$L_A(c\mathbf{v}) = A(c\mathbf{v}) = cA\mathbf{v} = cL_A(\mathbf{v}) \tag{4.15}$$

for all vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{K}$ and for all scalars $c \in \mathbb{K}$.

We call L_A the linear map **associated** with the matrix A. It can be easily shown that different matrices give rise to different associated maps. In other words, if two matrices give rise to the same linear map, then they are necessarily equal.

We've seen so far that any matrix in $\mathbb{K}^{m\times n}$ leads immediately to a linear map of $\mathbb{K}^n$ into $\mathbb{K}^m$. The opposite direction holds true, too.

Definition 4.10 (The matrix associated to a linear map). Let $L : \mathbb{K}^n \to \mathbb{K}^m$ be a linear map. Then there exists a *unique* matrix A such that $L = L_A$. The column vectors of A are the value of the linear map in the vectors of the canonical basis $\{e_1, \dots, e_n\}$ of $\mathbb{K}^n$. Precisely, for $j = 1, \dots, n$, the j^{th} column of A is found by applying L to the j^{th} standard basis vector

$$A = [L(e_1),\, \dots,\, L(e_n)] \in \mathbb{K}^{m\times n}. \tag{4.16}$$

Remark 4.2. The canonical basis for the space $\mathbb{C}^n$ is the same as for the space $\mathbb{R}^n$, so both definitions 4.9 and 4.10 are given for $\mathbb{K} = \mathbb{R}$ and $\mathbb{K} = \mathbb{R}$ in the same manner.

We've investigated so far the relation between matrices and linear maps of $\mathbb{K}^n$ into $\mathbb{K}^m$. A question arises naturally: do linear maps between arbitrary vector spaces over $\mathbb{K}$ have matrix representation? The answer, provided that the vector spaces are *finite*-dimensional, is *yes.* We will find out that, once the basis for each space are settled, the matrix representation of the map is unique.

Definition 4.11. Consider a finite-dimensional vector space V and let n be its dimension. Given a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V, any vector $\mathbf{v}$ in V admits the unique representation

$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \tag{4.17}$$

for suitable coefficients $a_1, \ldots, a_n \in \mathbb{K}$. Then we can consider the isomorphism

$$\Phi : \mathbb{K}^n \to V \tag{4.18}$$

defined by

$$(a_1, \ldots, a_n) \mapsto a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n \tag{4.19}$$

that maps the unique coordinates of any vector to the representation of the vector as a linear combination of the coordinates. We say that V is **isomorphic** to $\mathbb{K}^n$ under the map Φ and we write

$$V \cong \mathbb{K}^n. \tag{4.20}$$

Assume we are given two finite-dimensional vector spaces V and W over $\mathbb{K}$ and a linear map $L : V \to W$. Assume, moreover, that $\dim(V) = n$ and $\dim(W) = m$ and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be bases of V and W respective1y. Using the isomorphism Φ, we can identify $V \cong \mathbb{K}^n$ and $W \cong \mathbb{K}^m$ and interpret L as a linear map of $\mathbb{K}^n$ into $\mathbb{K}^m$, and thus we can associate a matrix with L. This matrix *strongly* depends on the choice of the basis. Precisely, it is the unique matrix A having the property that if we denote by $[\mathbf{x}]_{\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}}$ the (column) coordinate vector of a vector $\mathbf{v} \in V$, relative to the basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, then $A[\mathbf{x}]_{\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}}$ is the (column) coordinate vector of $L(\mathbf{x})$, relative to the basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$, that is

$$L(\mathbf{x})_{\{\mathbf{w}_1,\ldots,\mathbf{w}_n\}} = A[\mathbf{x}]_{\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}}. \tag{4.21}$$

We shall also write

$$A^{\{\mathbf{v}_1,\ldots,\mathbf{v}_n\}}_{\{\mathbf{w}_1,\ldots,\mathbf{w}_m\}}(L) \tag{4.22}$$

to denote that A is the matrix associated to the linear map L and to indicate the respective choices of the basis for V and W.

In short, the matrix carries all the essential information. If the basis is known, and the matrix is known, then the transformation of every vector is known. The coding of the information is simple. To transform a space to itself, one basis is enough. A transformation from one space to another requires a basis for each.

Definition 4.12. Let $I : \mathbb{R}^n \to \mathbb{R}^n$ be the identity mapping $v \mapsto I(v) := v$; it is a linear map:

i) $\forall v_1, v_2 \in \mathbb{R}^n \implies (v_1 + v_2) := v_1 + v_2 = I(v_1) + I(v_2)$;

ii) $\forall a \in \mathbb{K}, \forall v \in \mathbb{R}^n \implies I(av) := av = aI(v)$.

Another map which is easily shown to be linear is the null map, defined as follows:

$$0 : \mathbb{R}^n \to \mathbb{R}^n \text{ such that } v \mapsto 0(v) := 0.$$

With respect to the canonical basis of $\mathbb{R}^n$, the matrices associated to I and 0 are respectively:

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}, \quad 0 = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \in \mathbb{R}^{n \times n}$$

With a generic basis $\{v_1, \dots, v_2\}$ of $\mathbb{R}^n$, the identity matrix is $I = [I(v_1), \dots, I(v_n)] = [v_1, \dots, v_2]$.

Exercise 4.1. 1. Write down the matrix $A \in \mathbb{R}^{2\times 4}$ associated to the linear map $L : \mathbb{R}^4 \to \mathbb{R}^2$ such that:

$$L(e_1) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad L(e_2) = \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \quad L(e_3) = \begin{bmatrix} -5 \\ 4 \end{bmatrix}, \quad L(e_4) = \begin{bmatrix} 1 \\ 7 \end{bmatrix}.$$

Solution 4.1. Vectors obtained by applying L to the basis of $\mathbb{R}^2$ are the columns of A:

$$A = \begin{bmatrix} 2 & 3 & -5 & 1 \\ 1 & -1 & 4 & 7 \end{bmatrix}.$$

2. Write down the matrix $A \in \mathbb{R}^{3\times 2}$ associated to the linear map $L : \mathbb{R}^3 \to \mathbb{R}^2$ with respect to: the canonical bases:

$$\{e_1, e_2, e_3\} = \{[1,0,0]^T, [0,1,0]^T, [0,0,1]^T\}, \quad \{\widetilde{e_1}, \widetilde{e_2}\} = \{[1,0]^T, [0,1]^T\},$$

the canonical bases of $\mathbb{R}^3$ and $\mathbb{R}^2$. The vectors obtained applying L to $\{e_1, e_2, e_3\}$ are:

$$L(e_1) = \widetilde{e_1} + \widetilde{e_2}, \quad L(e_2) = 5\widetilde{e_1} + 5\widetilde{e_2}, \quad L(e_3) = 3\widetilde{e_1} - \widetilde{e_2}.$$

Solution 4.2. Representing $L(e_1), L(e_2)$ and $L(e_3)$ with respect to the canonical basis of $\mathbb{R}^2$ it follows that:

$$L(e_1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad L(e_2) = \begin{bmatrix} 5 \\ 5 \end{bmatrix}, \quad L(e_3) = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \implies A = \begin{bmatrix} 1 & 5 & 3 \\ 1 & 5 & -1 \end{bmatrix}.$$

4.4 Exercises

Exercise 4.2.

1. Find kernel, image, rank and nullity of the linear map $L : \mathbb{R}^3 \to \mathbb{R}$ such that $L(x, y, z) = 3x - 2y + z$.

   Solution 4.3. Since $L : V \simeq \mathbb{R}^3 \to W \simeq \mathbb{R}$, $\dim(V) = 3$. The kernel of F is:

   $$\text{Ker}(L) := \{v \in \mathbb{R}^3 \text{ such that } L(v) = 0\} = \{(x, y, z)^T : 3x - 2y + z = 0\}.$$

   The dimension of the kernel can be at most 3 since it a subspace of V. The image of F is:

   $$\text{Im}(L) = \{a \in \mathbb{R} \text{ such that } \exists (x, y, z)^T \in \mathbb{R}^3 : L(x, y, z)^T = 3x - 2y + z = a\}.$$

   Since the image of F is a subspace of W, its dimension can be only 0 (trivial) or 1. For example:

   $$(2, 1, 1)^T \implies 3 \cdot 2 - 2 \cdot 1 + 1 = 5 \in \mathbb{R};$$

   $\text{Im}(F)$ the canÕt be empty, so $\dim(\text{Im}(L)) = 1$. $3x - 2y + z = a$ is the equation of a plane trough the origin. The rank-nullity theorem holds:

   $$\dim(V) = 3 = \dim(\text{Ker}(F)) + \dim(\text{Im}(F)) = \dim(\text{Ker}(V)) + 1 \implies \dim(\text{Ker}(F)) = 2.$$

2. Consider the linear map $P : \mathbb{R}^3 \to \mathbb{R}^2$, $(x, y, z) \mapsto P(x, y, z) = (x, y)$. With the canonical basis of $\mathbb{R}^3$:

   $$P(e_1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad P(e_2) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad P(e_3) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \implies P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

   P is a projection which selects the first two components of a vector in $\mathbb{R}^3$. Write down the rank-nullity theorem.

Solution 4.4. According to the definition, the kernel of P is:

$$\text{Ker}(P) = \{(x, y, z)^T \in \mathbb{R}^3 \text{ such that } (x, y)^T = 0_{\mathbb{R}^2}\} = \{(0, 0, z)^T \in \mathbb{R}^3\}.$$

The kernel of P is the the vertical axis. The image of P is:

$$\text{Im}(P) = \{(x, y)^T \in \mathbb{R}^2 \text{ such that } \exists(x, y, z)^T \in \mathbb{R}^3 : P((x, y, z)^T) = (x, y)^T\} = \mathbb{R}^2.$$

We notice that the null element is included: $(x, y)^T = (0, 0)^T$. Since $\dim(\mathbb{R}^3) = 3$, it follows that:

$$\dim(\mathbb{R}^3) = 3 = \dim(\text{Ker}(P)) + \dim(\text{Im}(P)) = \dim(\text{Ker}(P)) + 2 \implies \text{null}(P) = \dim(\text{Ker}(P)) = 1.$$

We used the fact that since $\text{Im}(P) = \mathbb{R}^2$, $\text{rank}(P) = \dim(\text{Im}(P)) = 2$.

Definition 4.13. Given a matrix $A \in \mathbb{R}^{m\times n}$, we mutuate the following definitions:

- $\text{Ker}(A) := \{v \in \mathbb{R}^n \text{ such that } Av = 0\} \subseteq \mathbb{R}^n$;
- $\text{Im}(A) = \{w \in \mathbb{R}^m \text{ such that } \exists v \in \mathbb{R}^n \text{ such that } Av = W\} \subseteq \mathbb{R}^m$ (vector subspace).

The vectors $a_1, \ldots, a_n$, the columns of $A = [a_1, \ldots, a_n]$, span a vector subspace $\text{span}\{a_1, \ldots, a_n\} \subseteq \mathbb{R}^n$:

$$\dim(\text{span}\{a_1, \ldots, a_n\}) = \text{rank}(A) \leq n \rightarrow \text{column rank}.$$

The vectors $r_1, \ldots, r_n$, the rows of $A = [r_1, \ldots, r_n]^T$, span a vector subspace $\text{span}\{r_1, \ldots, r_n\} \subseteq \mathbb{R}^m$:

$$\dim(\text{span}\{r_1, \ldots, r_m\}) = \text{rank}(A) \leq m \rightarrow \text{row rank}.$$

It can be shown that the column and row rank of A are equal, so we simply call it $\text{rank}(A)$; $\text{rank}(A) \leq \min(m, n)$.

Chapter 5

Scalar products, norms and orthogonality

5.1 Scalar products on vector spaces

Definition 5.1 (Scalar product on a vector space)**.** Let V be a vector space over $\mathbb{R}$. A **scalar product** on V is a map

$$(\cdot,\cdot) : V \times V \to \mathbb{R} \tag{5.1}$$

satisfying the following properties

i) *bilinearity*: it is linear with respect to both arguments, which means that it is linear in the first argument when the second is kept constant

$$(c\mathbf{v}_1 + d\mathbf{v}_2, \mathbf{w}) = c(\mathbf{v}_1, \mathbf{w}) + d(\mathbf{v}_2, \mathbf{w}),\ \forall \mathbf{v}_1, \mathbf{v}_2, \mathbf{w} \in V \text{ and } \forall c, d \in \mathbb{R}$$

and vice versa

$$(\mathbf{v}, c\mathbf{w}_1 + d\mathbf{w}_2) = c(\mathbf{v}, \mathbf{w}_1) + d(\mathbf{v}, \mathbf{w}_2),\ \forall \mathbf{v}, \mathbf{w}_1, \mathbf{w}_2 \in V \text{ and } \forall c, d \in \mathbb{R};$$

ii) *symmetry*: $(\mathbf{v}_1, \mathbf{v}_2) = (\mathbf{v}_2, \mathbf{v}_1),\ \forall \mathbf{v}_1, \mathbf{v}_2 \in V$;

iii) *positive-definiteness*: $(\mathbf{v}, \mathbf{v}) \geq 0$, $\forall \mathbf{v} \in V$ and $(\mathbf{v}, \mathbf{v}) = 0$ if and only if $\mathbf{v} = \mathbf{0}_V$.

Remark 5.1. The scalar product is sometimes denoted by $\langle\cdot,\cdot\rangle$ and it is alternatively referred to as the **inner product**. A scalar product on a vector space V is also called a **symmetric positive-definite bilinear form**. The word "form" is used because the map (5.1) is scalar valued. We shall say that V is **endowed** with the scalar product $(\cdot,\cdot)$.

Remark 5.2. For $\mathbb{K} = \mathbb{C}$, the symmetry property becomes

ii) *hermitian*: $(\mathbf{v}_1, \mathbf{v}_2) = \overline{(\mathbf{v}_2, \mathbf{v}_1)}, \ \forall \mathbf{v}_1, \mathbf{v}_2 \in V$,

where $\overline{z}$ denotes the complex conjugate of the complex number $z \in \mathbb{C}$.

Example 5.1 (Standard Euclidean product in $\mathbb{R}^n$)**.** Let $\mathbf{x}$ and $\mathbf{y}$ denote vectors in $\mathbb{R}^n$. The standard **Euclidean product** in $\mathbb{R}^n$ is the map

$$(\cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R} \tag{5.2}$$

defined by

$$(\mathbf{x}, \mathbf{y}) := \mathbf{y}^T \mathbf{x} = \sum_{i=1}^{n} x_i y_i. \tag{5.3}$$

The standard **Euclidean product** in $\mathbb{C}^n$ is the map

$$(\cdot, \cdot) : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C} \tag{5.4}$$

defined by

$$(\mathbf{x}, \mathbf{y}) := \mathbf{y}^H \mathbf{x} = \sum_{i=1}^{n} x_i \overline{y}_i. \tag{5.5}$$

The scalar product in $\mathbb{R}^n$ (or $\mathbb{C}^n$) is often denoted by $\mathbf{x} \cdot \mathbf{y}$ and it is referred to as the **dot product** because of this notation.

Example 5.2. Let $V = C([a, b], \mathbb{R})$ be the set of all continuous functions defined on the interval $[a, b]$ of the real line, with $a < b$. Define for any $f, g \in V$

$$(f, g) := \int_a^b f(x) g(x)\, dx. \tag{5.6}$$

It can be easily verified that (5.6) defines a scalar product on V. The definition can be extended to $V = L^2([a, b])$ the set of square integrable functions on $[a, b]$.

Example 5.3. Let $\mathbb{R}^{m\times n}$ and $\mathbb{C}^{m\times n}$ be the vector spaces of $m \times n$ matrices with real and complex entries respectively. Define

$$(A, B) := tr(A^T B) \quad \text{and} \quad (A, B) := tr(A^H B) \tag{5.7}$$

for $A, B \in \mathbb{R}^{m\times n}$ and $\mathbb{C}^{m\times n}$ respectively, where the trace of an $n \times n$ matrix M is $tr(M) := \sum_{i=1}^{n} m_{ii}$. It can be shown that definitions (5.7) define scalar products in the space of $m \times n$ matrices and they are referred to as the **standard scalar products for matrices**. Observe that these reduce to the standard Euclidean scalar product for vectors in $\mathbb{R}^m$ when $n = 1$.

Definition 5.2 (Orthogonal vectors and orthogonal subspaces)**.** Let V be a real vector space endowed with a scalar product $(\cdot,\cdot)$. Two vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ are said to be **orthogonal**, or **perpendicular**, if $(\mathbf{v}_1, \mathbf{v}_2) = 0$ and this is denoted by writing $\mathbf{v}_1 \perp \mathbf{v}_2$.

Given a subset $S \subset V$, we define the **orthogonal subspace** of S by

$$S^{\perp} := \{\mathbf{w} \in V \text{ such that } (\mathbf{v}, \mathbf{w}) = 0 \ \forall \mathbf{v} \in S\}.$$

It can be easily proved that $S^{\perp}$ is a vector subspace of V.

Remark 5.3. Let $V = \mathbb{R}^n$ endowed with the Euclidean scalar product. Observe that the zero vector in $\mathbb{R}^n$ is orthogonal to every vector in $\mathbb{R}^n$ because $\mathbf{0}^T\mathbf{x} = \sum_{i=1}^{n} 0 \cdot x_i = 0$ for any $\mathbf{x}$ in $\mathbb{R}^n$.

Theorem 5.1. Let A be an $n \times n$ matrix over $\mathbb{R}$. The scalar product in $\mathbb{K}^n$ satisfies the following main property

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A^T\mathbf{y}), \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{5.8}$$

Let A be an $n \times n$ matrix over $\mathbb{C}$. Then

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A^H\mathbf{y}), \forall \mathbf{x}, \mathbf{y} \in \mathbb{C}^n. \tag{5.9}$$

Definition 5.3 (Orthogonal matrices)**.** An $n \times n$ matrix Q over $\mathbb{R}$ is called **orthogonal** if

$$QQ^T = Q^TQ = I. \tag{5.10}$$

Definition 5.4 (Unitary matrices)**.** An $n \times n$ matrix Q over $\mathbb{C}$ is called **unitary** if

$$QQ^H = Q^HQ = I. \tag{5.11}$$

Properties (5.8) and (5.9) have the important consequence that orthogonal (unitary) matrices preserve scalar products.

Theorem 5.2. Let Q be an orthogonal (unitary) matrix of order n over $\mathbb{K}$ ($\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$). Then

$$(Q\mathbf{x}, Q\mathbf{y}) = (\mathbf{x}, \mathbf{y}), \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{K}^n. \tag{5.12}$$

Proof. Assume that Q is orthogonal. Then property (5.8), the associative property of matrix multiplication and the definition of orthogonality (5.10) imply

$$(Q\mathbf{x}, Q\mathbf{y}) = (\mathbf{x}, Q^T(Q\mathbf{y})) = (\mathbf{x}, (Q^TQ)\mathbf{y}) = (\mathbf{x}, \mathbf{y}), \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{5.13}$$

The proof is the same for the complex case. □

Theorem 5.3. An $n \times n$ matrix A over $\mathbb{R}$ is orthogonal if and only if the row vectors of A form an orthonormal basis of $\mathbb{R}^n$, if and only if the column vectors of A form an orthonormal basis of $\mathbb{R}^n$.

5.2 Norms on vector spaces

Definition 5.5 (Norm on $\mathbb{K}$)**.** Let V be a vector space over $\mathbb{K}$. A **norm** on V is a map

$$\|\cdot\| : V \to \mathbb{K} \tag{5.14}$$

satisfying the following properties

i) *positive-definiteness*: $\|\mathbf{v}\| \geq 0,\ \forall \mathbf{v} \in V$ and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}_V$;

ii) $\|a\mathbf{v}\| = |a|\|\mathbf{v}\|\ \ \forall a \in \mathbb{K}, \forall v \in V$, where $|a|$ denotes the absolute value if a is a real number and the module if a is a complex number;

iii) *triangular inequality*: $\|\mathbf{v}_1 + \mathbf{v}_2\| \leq \|\mathbf{v}_1\| + \|\mathbf{v}_2\|\ \forall \mathbf{v}_1, \mathbf{v}_2 \in V$.

A vector space V with a norm $\|\cdot\|_V$ is a **normed vector space** and it is denoted by $(V, \|\cdot\|_V)$. If $\|\mathbf{v}\| = 1$, then $\mathbf{v}$ is called **unit vector**. If we divide a nonzero vector by its norm, we obtain a unit vector. This process is called **normalizing** the vector. Maps $|\cdot| : V \to \mathbb{K}$ that satisfy conditions ii) and iii), but are not positive definite are called **semi-norms**.

Example 5.4 (Hölder norms on $\mathbb{K}^n$)**.** Consider $(\mathbb{K}^n, \|\cdot\|_p)$, where $\|\cdot\|_p$ for $p \geq 1$ denotes the **Hölder norm** or the **p-norm** defined by

$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}. \tag{5.15}$$

For instance, for $p = 1$ we find the **grid norm**

$$\|\mathbf{x}\|_1 := \sum_{i=1}^{n} |x_i| \tag{5.16}$$

while the case $p = 2$ defines the **Euclidean norm** which is considered in Example 5.5.

As p goes to infinity, the limit of $\|\cdot\|_p$ exists and it is given by the **infinity (max) norm**

$$\|\mathbf{x}\|_\infty = \lim_{p\to\infty} \|\mathbf{x}\|_p = \max_{1\leq i\leq n} |x_i|. \tag{5.17}$$

For instance, if $\mathbf{x} = (-1, 0, 3)^T \in \mathbb{R}^3$, then $\|\mathbf{x}\|_1 = 4$ and $\|\mathbf{x}\|_\infty = 3$.

Example 5.5 (Norms induced by scalar products)**.** Let V be a vector space over $\mathbb{K}$. We shall say that the scalar product $(\cdot, \cdot)$ over V *induces* the norm $\|\cdot\|$ on V by letting

$$\|\mathbf{v}\|^2 := (\mathbf{v}, \mathbf{v}). \tag{5.18}$$

It can be proved that (5.18) verifies properties i), ii) and iii) in Definition 5.5.

An instance of induced norm is the 2-norm, which defines the **standard Euclidean norm**

$$\|\mathbf{x}\|_2^2 = \sum_{i=1}^{n} |x_i|^2 = \sum_{i=1}^{n} x_i^2 \tag{5.19}$$

that is *induced* by the standard Euclidean scalar product, since

$$(\mathbf{x}, \mathbf{x}) = \sum_{i=1}^{n} x_i^2 = \|\mathbf{x}\|_2^2. \tag{5.20}$$

Remark 5.4. Since each scalar product induces a norm by (5.18), the reverse question arises naturally. That is, given a vector norm $\|\cdot\|$ on a vector space V, can we define a corresponding scalar product on V such that $\sqrt{(\cdot,\cdot)} = \|\cdot\|$? If not, under what conditions will a given norm be generated by a scalar product? The answer to the former question is "no". In fact, given a vector norm $\|\cdot\|$ on a vector space V, there exists a scalar product on V such that $\sqrt{(\cdot,\cdot)} = \|\cdot\|$ if and only if the **parallelogram identity**

$$\|\mathbf{v}_1 + \mathbf{v}_2\| + \|\mathbf{v}_1 - \mathbf{v}_2\| = 2(\|\mathbf{v}_1\| + \|\mathbf{v}_2\|) \tag{5.21}$$

holds for all $\mathbf{v}_1, \mathbf{v}_1 \in V$. Take as an instance the standard Euclidean vector norm in $\mathbb{K}^n$. We already know that it is generated by the standard scalar product, so the parallelogram identity (5.21) must hold for the 2-norm. This is easily visualized geometrically in $\mathbb{R}^n$ for $n = 2, 3$. In fact, the parallelogram identity is so named because it expresses the fact that the sum of the squares of the diagonals in a parallelogram is twice the sum of the squares of the sides. However, it can be proved that the parallelogram identity (5.21) doesn't hold for $p \neq 2$. So except for the euclidean norm, the other p-norms cannot be generated by a scalar product.

Remark 5.5 (Geometric interpretation)**.** The Euclidean norm represents the **length** of vectors in $\mathbb{R}^n$. For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the **distance** between $\mathbf{u}$ and $\mathbf{v}$ is the length of the vector $\mathbf{u} - \mathbf{v}$, that is

$$\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|_2. \tag{5.22}$$

For instance, let $[a_1, a_2, a_3]$ and $[b_1, b_2, b_3]$ be the coordinates of respectively, $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^3$. Then

$$\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|_2 = \sqrt{(\mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v})} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2}. \tag{5.23}$$

Example 5.6 (Backward Triangle Inequality)**.** It can be easily shown by using the triangle inequality the following lower bound for the difference of vector norms

$$|\, \|\mathbf{v}_1\| - \|\mathbf{v}_2\| \,| \leq \|\mathbf{v}_1 - \mathbf{v}_2\| \; \forall \mathbf{v}_1, \mathbf{v}_2 \in V. \tag{5.24}$$

Theorem 5.4 (Phytagoras' theorem)**.** Let V be a real vector space and let $\|\cdot\|$ be the norm induced by the scalar product $(\cdot, \cdot)$ in V. Then

$$\|\mathbf{v}_1 + \mathbf{v}_2\|^2 = \|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2 \tag{5.25}$$

if and only if $\mathbf{v}_1 \perp \mathbf{v}_2$.

Proof. Let us measure the sum of two generic vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in V. We apply the definition of the norm induced the a scalar product and exploit the symmetry of the scalar product

$$\begin{aligned} \|\mathbf{v}_1 + \mathbf{v}_2\|^2 &= (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_1 + \mathbf{v}_2) \\ &= (\mathbf{v}_1, \mathbf{v}_1) + (\mathbf{v}_1, \mathbf{v}_2) + (\mathbf{v}_2, \mathbf{v}_1) + (\mathbf{v}_2, \mathbf{v}_2) \\ &= \|\mathbf{v}_1\|^2 + 2(\mathbf{v}_1, \mathbf{v}_2) + \|\mathbf{v}_2\|^2 \end{aligned} \tag{5.26}$$

We observe that (5.25) is satisfied if and only if $\mathbf{v}_1 \perp \mathbf{v}_2$. □

Remark 5.6. When V is a complex vector space space, then $\mathbf{v}_1 \perp \mathbf{v}_2$ if and only if $\|c\mathbf{v}_1 + d\mathbf{v}_2\|^2 = \|c\mathbf{v}_1\|^2 + \|d\mathbf{v}_2\|^2$ for all scalars c and d.

The following inequality is one of the most important inequalities in mathematics. It relates scalar products to the induced norms.

Theorem 5.5 (Cauchy-Schwartz inequality)**.** Let V be a vector space over $\mathbb{K}$ endowed with the scalar product $(\cdot, \cdot)$ and let $\|\cdot\|$ be the induced norm. Then

$$|(\mathbf{v}_1, \mathbf{v}_2)| \leq \|\mathbf{v}_1\| \|\mathbf{v}_2\| \quad \forall \mathbf{v}_1, \mathbf{v}_2 \in V \tag{5.27}$$

and $|(\mathbf{v}_1, \mathbf{v}_2)| = \|\mathbf{v}_1\| \|\mathbf{v}_2\|$ if and only if $\mathbf{v}_1 = c\mathbf{v}_2,$ for $c = (\mathbf{v}_1, \mathbf{v}_2)/\|\mathbf{v}_1\|$.

Definition 5.6 (Orthogonal bases)**.** Let V be a vector space over $\mathbb{K}$ endowed with the scalar product $(\cdot, \cdot)$ which induces the norm $\|\cdot\|$. A basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of V is called an **orthogonal basis** if

$$\mathbf{v}_i \perp \mathbf{v}_j, \; \forall i, j = 1, \dots, n \text{ and } i \neq j. \tag{5.28}$$

If in addition $\|\mathbf{v}_i\| = 1 \; \forall i = 1, \dots, n$, the basis is called an **orthonormal basis**.

Remark 5.7. In general, a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of n vectors in V is called an orthonormal set if $\mathbf{v}_i \perp \mathbf{v}_j,\ \forall i, j = 1, \ldots, n$ and $i \neq j$ and $\|\mathbf{v}_i\| = 1\ \forall i = 1, \ldots, n$. It can be proved that any orthonormal set is linearly independent and hence every orthonormal set of n vectors in an n-dimensional space V is an orthonormal basis for V.

Example 5.7. The standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$ is an orthonormal basis.

Remark 5.8. It is important to know that if V is a finite dimensional vector space, then there always exists an orthonormal basis for V. It can be constructed through the well-known Graham-Schmidt procedure. If we represent vectors of V in coordinates with respect to this basis, for instance $\mathbf{u} = [a_1, a_2, \ldots, a_n]$ and $\mathbf{v} = [b_1, b_2, \ldots, b_n]$, then

$$(\mathbf{u}, \mathbf{v}) = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n. \tag{5.29}$$

Exercise 5.1. Determine which of the following vectors in $\mathbb{C}^3$ forms an orthonormal basis of $\mathbb{C}^3$

$$\mathbf{w} = \left[\frac{i}{\sqrt{3}}, \frac{i}{\sqrt{3}}, \frac{i}{\sqrt{3}}\right]^T,\ \mathbf{x} = \left[\frac{-2i}{\sqrt{6}}, \frac{i}{\sqrt{6}}, \frac{i}{\sqrt{6}}\right]^T,\ \mathbf{y} = \left[\frac{i}{\sqrt{6}}, \frac{i}{\sqrt{6}}, \frac{-2i}{\sqrt{6}}\right]^T \tag{5.30}$$

and

$$\mathbf{z} = \left[0, \frac{-i}{\sqrt{2}}, \frac{i}{\sqrt{2}}\right]^T. \tag{5.31}$$

Solution 5.1. Vectors in an orthonormal basis must be unit vectors so first compute their norms to see if we can exclude anyone and we find out that

$$\|\mathbf{w}\|_2^2 = \|\mathbf{x}\|_2^2 = \|\mathbf{y}\|_2^2 = \|\mathbf{z}\|_2^2 = 1. \tag{5.32}$$

We can't exclude anyone, so we investigate their mutual orthogonality and we find out that

$$(\mathbf{w}, \mathbf{x}) = (\mathbf{w}, \mathbf{y}) = (\mathbf{w}, \mathbf{z}) = (\mathbf{x}, \mathbf{z}) = 0, \quad (\mathbf{x}, \mathbf{y}) \neq 0 \text{ and } (\mathbf{y}, \mathbf{z}) \neq 0. \tag{5.33}$$

We see that the only candidate for a basis is $\{\mathbf{w}, \mathbf{x}, \mathbf{z}\}$. We can easily prove that these vectors are linearly independent. Since $\dim(\mathbb{C}^3) = 3$, we conclude that $\{\mathbf{w}, \mathbf{x}, \mathbf{z}\}$ form an orthonormal basis of $\mathbb{C}^3$.

Definition 5.7 (Equivalent norms)**.** Two norms $||\cdot||$ and $|||\cdot|||$ on V are equivalent if there exist two positive constants c and C such that:

$$c|||\mathbf{v}||| \leq ||\mathbf{v}|| \leq C|||\mathbf{v}||| \quad \forall \mathbf{v} \in V. \tag{5.34}$$

Theorem 5.6. All norms in finite dimensional vector spaces are equivalent.

For instance, all norms in $\mathbb{R}^n$ are equivalent. For example, it can be shown that

$$n^{-1/2}\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_\infty \leq 1\|\mathbf{x}\|_2, \tag{5.35}$$

where the constants are $c = n^{-1/2}$ and $C = 1$.

5.3 Matrix norms

Matrix norms can be used to provide a measure of distance on the space of matrices. They play an important role in the analysis of matrix algorithms. For instance, they are useful for assessing the accuracy of computations or for measuring progress during an iteration.

Definition 5.8. A **matrix norm** is a map

$$\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R} \tag{5.36}$$

satisfying the following properties

i) $\|A\| \geq 0$, $\forall A \in \mathbb{C}^{m\times n}$ and $\|A\| = 0$ if and only if $A = O$ (*positive-definiteness*);

ii) $\|aA\| = |a|\|A\| \;\; \forall a \in \mathbb{C}, \forall A \in \mathbb{C}^{m\times n}$;

iii) $\|A + B\| \leq \|A\| + \|B\| \quad \forall A, B \in \mathbb{C}^{m\times n}$ (*triangular inequality*);

The most frequently used matrix norms in numerical linear algebra are the Frobenius norm and the matrix p-norms induced by the vector p-norms.

Example 5.8 (Frobenius norm)**.** The Frobenius norm is defined by the equations

$$\|A\|_F := \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^2} = \sqrt{\mathrm{tr}(AA^H)}. \tag{5.37}$$

Recall that tr($\bullet$) denotes the trace of a matrix M and it is defined by $\mathrm{tr}(A) := \sum_{i=1}^{n} m_{ii}$. By trivially applying definition (5.37) one finds out that the Frobenius norm of the identity matrix of order n concides with its trace

$$||I_n||_F^2 = 1^2 + \cdots + 1^2 = n = \mathrm{tr}(I_n). \tag{5.38}$$

Example 5.9 (The matrix p-norm)**.** The vector p-norm on $\mathbb{C}^n$ induces (or generates) the matrix p-norm on $\mathbb{C}^{m\times n}$ by defining

$$\|A\|_p := \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \quad \forall A \in \mathbb{C}^{m\times n} \text{ and } \forall \mathbf{x} \in \mathbb{C}^n. \tag{5.39}$$

It is clear that $\|A\|_p = \max_{\|\mathbf{x}\|_p=1} \|A\mathbf{x}\|_p$. It is important to notice that (5.37) defines a family of norms, depending on the space of matrices where they are defined. For instance, the 2-norm on $\mathbb{C}^{3\times 2}$ is different from the 2-norm on $\mathbb{C}^{6\times 5}$.

If $A \in \mathbb{C}^{m\times n}$, the 1-norm is the maximum of the absolute column sums

$$\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}| \tag{5.40}$$

and the ∞-norm is the maximum of the absolute row sums

$$\|A\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|. \tag{5.41}$$

Example 5.10 (Matrix norms induced by vector norms)**.** More generically, any vector norm $\|\bullet\|$ on $\mathbb{C}^n$ induces (or generates) a matrix norm $\|\bullet\|$ on $\mathbb{C}^{m\times n}$ by setting

$$\|A\| := \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \quad \forall A \in \mathbb{C}^{m\times n} \text{ and } \forall \mathbf{x} \in \mathbb{C}^n. \tag{5.42}$$

The norm of A according to (5.42) measures the largest amount by which any vector is amplified by matrix multiplication. Trivially, (5.42) implies that

$$\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\| \quad \forall A \in \mathbb{C}^{m\times n} \text{ and } \forall \mathbf{x} \in \mathbb{C}^n. \tag{5.43}$$

In other words, we can equivalently say that $\|A\|$ bounds the "amplifying power" of the matrix A, for all vectors $\mathbf{x}$. It is evident that $\|A\| = \max_{\|\mathbf{x}\|=1} \|A\mathbf{x}\|$. This suggests that the induced matrix norm $\|A\|$ represents the maximum extent to which a vector on the unit sphere $\mathcal{S} = \{\mathbf{x} \in \mathbb{C}^n \mid \|\mathbf{x}\| = 1\}$ can be stretched by the matrix A.

When A is nonsingular

$$\min_{\|\mathbf{x}\|=1} \|A\mathbf{x}\| = \frac{1}{\|A^{-1}\|}, \tag{5.44}$$

suggesting that $1/\|A^{-1}\|$ measures the extent to which a nonsingular matrix A can shrink vectors on $\mathcal{S}$.

Definition 5.9. A matrix norm $\|\bullet\|$ is **compatible** (or **consistent**) with a vector norm $\| \bullet \|$ if

$$\|A\mathbf{x}\| \leq \|A\|\|\mathbf{x}\| \quad \forall A \in \mathbb{C}^{m\times n} \quad \text{and} \quad \forall \mathbf{x} \in \mathbb{C}^n. \tag{5.45}$$

The p-norms satisfy the compatibility

$$\|A\mathbf{x}\|_p \leq \|A\|_p\|\mathbf{x}\|_p \quad \forall A \in \mathbb{C}^{m\times n} \quad \text{and} \quad \forall \mathbf{x} \in \mathbb{C}^n. \tag{5.46}$$

More generically, it is apparent by (5.43) that an induced matrix norm is compatible with its underlying vector norm. As another important instance, it can be proved that the Frobenius matrix norm $\| \bullet \|_F$ and the Euclidean vector norm $\| \bullet \|_2$ are compatible

$$\|A\mathbf{x}\|_2 \leq \|A\|_F\|\mathbf{x}\|_2 \quad \forall A \in \mathbb{C}^{m\times n} \quad \text{and} \quad \forall \mathbf{x} \in \mathbb{C}^n. \tag{5.47}$$

.

Definition 5.10. A matrix norm is called **sub-multiplicative** if

$$\|AB\| \leq \|A\|\|B\| \quad \forall A \in \mathbb{C}^{m\times n} \quad \text{and} \quad \forall B \in \mathbb{C}^{n\times s}. \tag{5.48}$$

Not all matrix norms satisfy this property. An instance of a matrix norm which is not sub-multiplicative is given by

$$||A||_\Delta := \max_{i,j=1,\ldots,n} |a_{ij}|, \quad \forall A \in \mathbb{C}^{m\times n}. \tag{5.49}$$

Indeed, for

$$A = B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{5.50}$$

we find that

$$\|A\|_\Delta = \|B\|_\Delta = 1, \tag{5.51}$$

but for

$$AB = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \tag{5.52}$$

we have that

$$\|A\|_\Delta\|B\|_\Delta = 1 < 2 = \|AB\|_\Delta. \tag{5.53}$$

Nevertheless, we mostly work with norms that satisfy (5.48). Many textbooks embed the sub-multiplicative property in the definition of matrix norms.

Theorem 5.6 implies that all norms in $\mathbb{C}^{m\times n}$ are equivalent. Some equivalence relationships hold as follows

$$\|A\|_2 \le \|A\|_F \le \sqrt{\min\{m,n\}}\|A\|_2, \tag{5.54}$$
$$\frac{1}{\sqrt{n}}\|A\|_\infty \le \|A\|_2 \le \sqrt{m}\|A\|_\infty, \tag{5.55}$$
$$\frac{1}{\sqrt{m}}\|A\|_1 \le \|A\|_2 \le \sqrt{n}\|A\|_1. \tag{5.56}$$

The efficacy of the matrix 1- and ∞-norms is that they are easy computations ($\mathcal{O}(n^2)$), while the calculation of the 2-norm is a more involved computation. Finally, in Chapter 6, we would like to stress out the theoretical importance of the 2-norm, along with some additional properties.

5.4 Exercices

Exercise 5.2. Consider the matrix

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & 5 \\ 4 & -1 & 3 \end{bmatrix}. \tag{5.57}$$

Compute the 1-, the ∞- and the Frobenius norm of A.

Solution 5.2. The 1-norm of A is

$$\begin{aligned} \|A\|_1 &= \max_{1\le j\le 3} \sum_{i=1}^{3} |a_{ij}| \\ &= \max\{1+2+4, 0+3+1, 1+5+3\} \\ &= \max\{7,4,9\} \\ &= 9. \end{aligned} \tag{5.58}$$

The ∞-norm of A is

$$\begin{aligned} \|A\|_\infty &= \max_{1\le i\le 3} \sum_{j=1}^{3} |a_{ij}| \\ &= \max\{1+0+1, 2+3+5, 4+1+3\} \\ &= \max\{2,10,8\} \\ &= 10. \end{aligned} \tag{5.59}$$

The Frobenius norm of A is

$$\begin{aligned} \|A\|_F &:= \sqrt{\sum_{i,j=1}^{3} |a_{ij}|^2} \\ &= \sqrt{1+1+4+9+25+16+1+9} \\ &= \sqrt{66}. \end{aligned} \tag{5.60}$$

Chapter 6

Eigenvalues and eigenvectors

6.1 Eigenvalues and eigenvectors

Definition 6.1 (Eigenvalues and eigenvectors of a linear map)**.** Let V be a vector space over $\mathbb{K}$ ($\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$) and let $L : V \to V$ be a linear map. A *nonzero* vector $\mathbf{v} \in V$ is called an **eigenvector** or an **eigenfunction** of L if there exists a scalar $\lambda \in \mathbb{K}$ such that

$$L\mathbf{v} = \lambda\mathbf{v}. \tag{6.1}$$

The scalar λ is called an **eigenvalue** of L corresponding to the eigenvector $\mathbf{v}$. In other words, L acts on eigenvectors as multiplication by scalars (the corresponding eigenvalues).

Remark 6.1. Note that the definition of eigenvector requires that it is a nonzero vector. In fact, if $\mathbf{v} = \mathbf{0}$ were allowed, then any scalar λ would be an eigenvalue since $L\mathbf{0} = \lambda\mathbf{0}$ holds for any $\lambda \in \mathbb{K}$. On the other hand, we can have $\lambda = 0_{\mathbb{K}}$, and $\mathbf{v} \neq \mathbf{0}$.

Definition 6.2 (Eigenvalues and eigenvectors of a matrix)**.** Let A be an $n \times n$ matrix over $\mathbb{K}$. An eigenvector of A is an eigenvector of the associated linear map $L_A : \mathbb{K}^n \to \mathbb{K}^n$, defined by $L_A\mathbf{v} = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{K}^n$. Thus an eigenvector of A is a nonzero vector $\mathbf{v} \in \mathbb{K}^n$ for which there exists $\lambda \in \mathbb{K}$ such that $A\mathbf{v} = \lambda\mathbf{v}$.

Geometrically, (6.1) means that eigenvectors are those vectors that experience only changes in magnitude or sign under the action of L. Specifically, the eigenvalue λ is simply the amount of "stretch" or "shrink" to which the eigenvector $\mathbf{v}$ is subjected when transformed by L. For instance, $L\mathbf{v} = 2\mathbf{v}$ says that $\mathbf{v}$ is stretched by factor 2 under transformation by L. Similarly, $L\mathbf{v} = (-1/3)\mathbf{v}$ says that L reverses $\mathbf{v}$ and shrinks it by a factor 1/3.

The prefix eigen- is adopted from the German word "eigen", which means "owned by" or "peculiar", "specific", "characteristic" to. It was the mathematician David Hilbert who introduced the terms *Eigenwert* and *Eigenfunktion*. Eigenvalues and eigenvectors are equivalently called **characteristic values** and **characteristic vectors**, **proper values** and **proper vectors**, or **latent values** and **latent vectors**. The couple $(\lambda, \mathbf{v})$ is also called **eigenpair**.

Eigenvectors are never unique. If $\mathbf{v}$ is an eigenvector for L with associated eigenvalue λ, then so is $c\mathbf{v}$, for any $c \in \mathbb{K}$ with $c \neq 0$. Indeed,

$$L(c\mathbf{v}) = cL\mathbf{v} = c(\lambda\mathbf{v}) = \lambda(c\mathbf{v}). \tag{6.2}$$

More generally, let

$$\begin{aligned} E_\lambda &:= \{\mathbf{v} \in V | L\mathbf{v} = \lambda\mathbf{v}\} \\ &:= \{\text{eigenvectors associated with } \lambda\} \cup \{\mathbf{0}_V\}. \end{aligned} \tag{6.3}$$

E_λ is called the **eigenspace** of L corresponding to λ. By observing that (6.1) can we rewritten as $(L - \lambda I_V)\mathbf{v} = \mathbf{0}$, we deduce that $E_\lambda = \mathrm{Ker}(L - \lambda I_V)$.

Theorem 6.1. Let V be a finite dimensional vector space and let $L : V \to V$ be a linear map. Then $\lambda \in \mathbb{K}$ is an eigenvalue of L if and only if $E_\lambda \neq \{\mathbf{0}_V\}$ if and only if $L - \lambda I$ is not invertible.

For matrices $A \in \mathbb{K}^{n \times n}$, the eigenspace associated to an eigenvalue λ is by definition the eigenspace of L_A corresponding to λ, hence $E_\lambda = \mathrm{Ker}(A - \lambda I_n)$. Similarly to Theorem 6.1, we state the following important theorem.

Theorem 6.2. Let A be an $n \times n$ matrix over $\mathbb{K}$. Then $\lambda \in \mathbb{K}$ is an eigenvalue of A if and only if $E_\lambda \neq \{\mathbf{0}_{\mathbb{K}^n}\}$ if and only if the matrix $A - \lambda I_n$ is not invertible if and only if $\det(A - \lambda I_n) = 0$.

The following theorem is an important result that states that to different eigenvalues correspond linearly independent eigenvectors.

Theorem 6.3. Let V be a vector space over $\mathbb{K}$ and let $L : V \to V$ be a linear map. Let $\lambda_1, \ldots, \lambda_k$ be eigenvalues of L, and let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be associated eigenvectors. If $\lambda_i \neq \lambda_j$ for any $i \neq j$, then the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are linearly independent.

Corollary 6.1. Let V be a finite dimensional vector space of dimension n and let $L : V \to V$ be a linear map having n eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ whose eigenvalues $\lambda_1, \ldots, \lambda_n$ are distinct. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V.

An important consequence of Theorem 6.3 is that any $n \times n$ matrix A over $\mathbb{K}$ can have at most n distinct eigenvalues in $\mathbb{K}$. Indeed, if A had $m > n$ distinct eigenvalues, then the associated eigenvectors would be a set of m linearly independent vectors of $\mathbb{K}^n$, which is impossible (cf. Theorem 2.1).

Next, we define the important notions of spectrum and spectral radius.

Definition 6.3 (Spectrum of a linear map)**.** Let V be a vector space over $\mathbb{K}$ and let $L : V \to V$ be a linear map. The set of distinct eigenvalues of L is called the **spectrum** of L and it is denoted by $\sigma(L)$.

By definition, the spectrum of an $n \times n$ matrix A is the spectrum of L_A and it is denoted by $\sigma(A)$. For instance, the eigenvalues of $A = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$ are 0 and 4 and therefore $\sigma(A) = \{0, 4\}$.

Definition 6.4 (Spectral radius of a matrix)**.** Let A be an $n \times n$ matrix over $\mathbb{K}$. The **spectral radius** of A is defined by

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|. \tag{6.4}$$

We observe that for any induced matrix norm $\| \bullet \|$ it holds

$$|\lambda| \leq \|A\| \quad \text{for all } \lambda \in \sigma(A) \tag{6.5}$$

Indeed, given an eigenpair $(\lambda, \mathbf{v})$ where $\mathbf{v} \neq \mathbf{0}$, $\lambda \mathbf{v} = A\mathbf{v}$ implies $|\lambda| \|\mathbf{v}\| = \|A\mathbf{v}\| \leq \|A\| \|\mathbf{v}\|$, from which we deduce (6.5). Equation (6.5) leads to the crude (but cheap) upper bound on $\rho(A)$

$$\rho(A) \leq \|A\| \quad \text{for any } \| \bullet \|. \tag{6.6}$$

6.2 Computation of eigenpairs

We address the following questions. If V is a vector space over $\mathbb{K}$, and $L : V \to V$ is a linear map, how can we find out the set of eigenvalues of L? Once the latest are at our disposal, how can we find out the associated eigenvectors (or eigenspaces)?

Definition 6.5. Given an $m \times n$ matrix A over $\mathbb{K}$, the **characteristic polynomial** associated with A is defined by

$$p_A(\lambda) = \det(A - \lambda I_n). \tag{6.7}$$

Notice that by expanding according to the first column, we find

$$\begin{aligned} p_A(\lambda) &= \det(A - \lambda I) \\ &= \begin{vmatrix} a_{11} - \lambda & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} - \lambda \end{vmatrix} \\ &= (a_{11} - \lambda) \cdots (a_{nn} - \lambda) + \cdots \\ &= (-1)^n \lambda^n + \text{low order terms in } \lambda, \end{aligned} \tag{6.8}$$

we deduce that $p_A(\lambda)$ is a polynomial of degree n in the variable λ, whose leading term is $(-1)^n \lambda^n$.

By Theorem 6.2, the scalar λ is an eigenvalue for any given $m \times n$ matrix A if and only if $p_A(\lambda) = 0$. The equation $p_A(\lambda) = 0$ is called the **characteristic equation** of the matrix A. Determinants give us a direct way of computing eigenvalues of a matrix, since the latest are are the solutions of the characteristic equation or, equivalently, the roots of the characteristic polynomial. Clearly, this is possible *whenever* we can determine explicitly the roots of $p_A(\lambda)$. Sometimes this is an easy task but in the majority of cases it is a formidable task.

Example 6.1. Find the eigenvalues and the associated eigenspaces for

$$A = \begin{pmatrix} 1 & 9 \\ 1 & 1 \end{pmatrix}. \tag{6.9}$$

We have that

$$A - \lambda I = \begin{pmatrix} 1 - \lambda & 9 \\ 1 & 1 - \lambda \end{pmatrix} \quad \text{and} \quad p_\lambda(A) = (1 - \lambda)^2 - 9 = \lambda^2 - 2\lambda - 8. \tag{6.10}$$

This factors as $p_A(\lambda) = (\lambda - 4)(\lambda + 2)$, so there are two eigenvalues: $\lambda_1 = 4$, and $\lambda_2 = -2$.

Once the eigenvalues are known, it is straightforward to compute the corresponding eigenvectors and eigenspaces. Indeed, in order to find the eigenvectors corresponding to λ_i for $i = 1, 2$ we apply Definition (6.1) and we seek for a non-trivial solution to the corresponding homogeneous equation $(A - \lambda_i I)\mathbf{v} = \mathbf{0}$.

For $\lambda_1 = 4$, let $\mathbf{v}_1 = (v_{1,1}, v_{1,2})^T$ denote a corresponding eigenvector, which satisfies

$$\begin{pmatrix} 1 - \lambda & 9 \\ 1 & 1 - \lambda \end{pmatrix} \begin{pmatrix} v_{1,1} \\ v_{1,2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{6.11}$$

This leads to a linear system of equations

$$\begin{cases} -3v_{1,1} + 3v_{1,2} = 0 \\ 3v_{1,1} - 3v_{1,2} = 0 \end{cases} \tag{6.12}$$

Notice that the equations are multiples of each other, so it is sufficient to solve one of them. The general solution of each of them consists of all vectors $\mathbf{v}$ such that

$$\mathbf{v} = \begin{pmatrix} v_{1,1} \\ v_{1,2} \end{pmatrix} = c \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{where } c \text{ is arbitrary.} \tag{6.13}$$

Eigenvectors are obtained for all $c \neq 0$. For instance, by choosing $c = -3$, we get an eigenvector $\mathbf{v}_1 = \begin{pmatrix} -3 \\ -3 \end{pmatrix}$. The set of all eigenvectors is a line with the origin missing and the eigenspace for $\lambda_1 = 4$ is

$$E_4 := \left\{ c \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ where } c \text{ is arbitrary} \right\} \cup \{\mathbf{0}\}. \tag{6.14}$$

It has dimension one and a possible basis is given by $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Note that any non-zero scalar multiples of this vector is also a basis.

We can proceed similarly for $\lambda_2 = -2$.

The following example shows that the eigenvalues strongly depend on the underlying field $\mathbb{K}$.

Example 6.2. Consider the 2×2 matrix A with real entries (over $\mathbb{R}$)

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{6.15}$$

We have that

$$A - \lambda I = \begin{pmatrix} -\lambda & 1 \\ -1 & -\lambda \end{pmatrix} \quad \text{and} \quad p_\lambda(A) = \lambda^2 + 1. \tag{6.16}$$

The characteristic equation $\lambda^2 + 1 = 0$ has no solutions over the field of real numbers, so there are no eigenvalues or eigenvectors. However, if we consider A defined over $\mathbb{C}$, then there are two eigenvalues i and $-i$, with corresponding eigenvectors the scalar multiples of the vectors $\begin{pmatrix} -1 \\ i \end{pmatrix}$ and $\begin{pmatrix} 1 \\ i \end{pmatrix}$ respectively.

Let's complete the discussion by recalling a powerful result given by fundamental theorem of algebra, which states that every polynomial of degree n in the domain of complex numbers has exactly n roots. Some of these roots may be complex numbers (even if all the coefficients are real), and some roots may be repeated. Therefore, altogether any $n \times n$ matrix A over $\mathbb{C}$ has exactly n eigenvalues. Some eigenvalues might lie in $\mathbb{C}$, while others might be repeated. In particular, if the entries of A lie in $\mathbb{R} \subset \mathbb{C}$, complex eigenvalues occur in conjugate pairs, i.e. if $\lambda \in \sigma(A)$, then $\bar{\lambda} \in \sigma(A)$. This is a consequence of the fact that the roots of a polynomial with real coefficients occur in conjugate pairs.

Definition 6.6 (Multiplicities)**.** Let $\lambda \in \sigma(A)$ be an eigenvalue of A. The number of times λ is repeated as a root of $p_\lambda(A)$ is called the **algebraic multiplicity** of λ and it is denoted by $m_a(\lambda)$. Eigenvalues that occur only once as roots of $p_\lambda(A)$ have $m_a(\lambda) = 1$. These are called **simple eigenvalues** of A. The **geometric multiplicity** of λ is $\dim(E_\lambda) = \dim(\mathrm{Ker}(A - \lambda I))$ and it is denoted by $m_g(\lambda)$. In other words, the geometric multiplicity of λ is the maximal number of linearly independent eigenvectors corresponding to λ. Eigenvectors for which $m_g(\lambda) = m_a(\lambda)$ are called **semisimple eigenvalues** of A.

Theorem 6.4. Let A be an $n \times n$ matrix over $\mathbb{C}$ and let $\lambda \in \sigma(A)$. Then,

$$1 \leq m_g(\lambda) \leq m_a(\lambda). \tag{6.17}$$

Let's see now some examples of computations of particularly favorable matrices.

Example 6.3. The eigenvalues of a diagonal matrix are its diagonal entries. For example, the matrix

$$A = \begin{pmatrix} -2 & 0 \\ 0 & 3 \end{pmatrix} \tag{6.18}$$

has eigenvalues $\lambda_1 = -2$ with corresponding eigenspace

$$E_{-2} := \left\{ c \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ where } c \text{ is arbitrary} \right\} \cup \{\mathbf{0}\} \tag{6.19}$$

and $\lambda_2 = 3$ with corresponding eigenspace

$$E_3 := \left\{ c \begin{pmatrix} 0 \\ 1 \end{pmatrix} \text{ where } c \text{ is arbitrary} \right\} \cup \{\mathbf{0}\}. \tag{6.20}$$

Example 6.4. The only possible values for the eigenvalues of a projection matrix are 0 and 1. For example, the matrix

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{6.21}$$

has eigenvalues $\lambda_1 = 1$ with corresponding eigenspace

$$E_1 := \left\{ c \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ where } c \text{ is arbitrary} \right\} \cup \{\mathbf{0}\} \tag{6.22}$$

and $\lambda_2 = 0$ with corresponding eigenspace

$$E_0 := \left\{ c \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ where } c \text{ is arbitrary} \right\} \cup \{\mathbf{0}\}. \tag{6.23}$$

Everytime $\lambda = 1$, an associated eigenvector is projected into itself, and everytime $\lambda = 0$, an associated eigenvector is projected to the zero vector. Like every other scalar, zero might or might be not an eigenvalue and there is nothing exceptional about that. There is an information that we can deduce if zero is an eigenvalue and it is that A is singular (not invertible), i.e. its determinant is zero. Invertible matrices must have all eigenvalues different from zero.

Example 6.5. Every triangular matrix has the eigenvalues sitting along the main diagonal. Indeed, consider the 3×3 matrix

$$A = \begin{pmatrix} 1 & 9 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & -4 \end{pmatrix} . \tag{6.24}$$

Since the determinant of any triangular matrix is the product of the diagonal entries, we have that

$$p_A(\lambda) = \det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 9 & 3 \\ 0 & 2-\lambda & -1 \\ 0 & 0 & -4-\lambda \end{vmatrix} = (1-\lambda)(2-\lambda)(-4-\lambda) \tag{6.25}$$

is zero if and only if $\lambda = 1, 2, -4$, which give the diagonal entries of A.

6.3 Diagonalization of a matrix

Definition 6.7 (Similar matrices)**.** Two $n \times n$ matrices A and B over $\mathbb{K}$ are called **similar** if there exists an $n \times n$ invertible matrix P such that $B = P^{-1}AP$. Going from one to the other is called a **similarity transformation**.

Similar matrices have the same characteristic equation, hence the same eigenvalues. Indeed, by exploiting the properties of the determinant one can easily deduce that $\det(B) = \det(P^{-1}AP)$.

In particular, we recall the different matrices corresponding to a linear map are all similar,(spiega fai riferimento)therefore they all have the same characteristic equation. In particular, this implies that the eigenvalues of a linear map are independent on the particular choice of the basis.

Definition 6.8. An $n \times n$ matrix over $\mathbb{K}$ is **diagonalizable** if it is similar to a diagonal matrix.

There is a strong connection between diagonalizable matrices and eigenvectors. More specifically, suppose that A has n linearly independent eigenvectors. Let S be the square matrix whose columns are given by the eigenvectors of A. Then, it turns out that the matrix $S^{-1}AS$ is a diagonal matrix Λ whose diagonal entries are the eigenvalues of A:

$$S^{-1}AS = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \tag{6.26}$$

We will refer to S as the **eigenvector matrix** and Λ the **eigenvalue matrix**. The eigenvector matrix S converts the matrix A into a diagonal matrix (its eigenvalue matrix Λ).

Remark 6.2. If the matrix A has n distinct eigenvalues, then the corresponding n eigenvectors are automatically linearly independent (cf. Thm 6.3). Consequently, any $n \times n$ matrix with n distinct eigenvalues can by diagonalized. The converse is not true. Matrices that have repeated eigenvalues may be a case of not diagonalizable matrices, but in general it is not said–it only depends on whether the corresponding n eigenvectors are linearly independent or not. An obvious counterexample is given by the identity matrix of order n, which has only a single eigenvalue, the scalar $1_{\mathbb{K}}$, repeated n times, but it is diagonal already.

Remark 6.3. The only possible matrices S that diagonalize a given matrix A are those whose columns are eigenvectors of A. Other choices of S will not produce a diagonal matrix from $S^{-1}AS$.

Remark 6.4. The eigenvector matrix S in representation (6.26) is not unique. Indeed, each eigenvector can be multiplied by a nonzero scalar and it remains an eigenvector. If we pre-multiply the columns of S by any nonzero constant, we will get a new valid S for which $S^{-1}AS = \Lambda$.

Remark 6.5. Not all matrices are diagonalizable since not all matrices posses n distinct eigenvectors and we cannot construct S in those cases. These are called **defective matrices**.

Symmetric matrices defined over $\mathbb{R}$ enjoy some nicer properties. Their eigenvalues are all real, eigenvectors corresponding to distinct eigenvalues are orthogonal, and those matrices can be diagonalized by real orthogonal matrices.

Theorem 6.5. Let A be a real symmetric matrix. Then A has an eigenvalue in $\mathbb{R}$, and all complex eigenvalues of A lie in $\mathbb{R}$.

Theorem 6.6. Let A be a real symmetric $n \times n$ matrix, where $\mathbb{R}^n$ is endowed with the Euclidean scalar product. Let λ_1, λ_2 be two distinct eigenvalues of A, with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2$, then $\mathbf{v}_1 \perp \mathbf{v}_2$.

Theorem 6.7. Let A be a real symmetric $n \times n$ matrix. Then there exists a real orthogonal matrix Q with $Q^{-1}AQ = Q^TAQ = \Lambda$.

Bibliography

[1] Serge Lang. *Introduction to linear algebra*. Springer Science & Business Media, 2012.

[2] Gene H Golub and Charles F Van Loan. *Matrix computations*, volume 3. JHU press, 2012.

[3] David C Lay. *Linear algebra and its applications*. Addison-Wesley,, 2012.

[4] Carl D Meyer. *Matrix analysis and applied linear algebra*, volume 71. Siam, 2000.

[5] Gilbert Strang. Linear algebra and its applications. 2006. *Thomson, Brooks/Cole*.

[6] Alfio Quarteroni, Riccardo Sacco, and Fausto Saleri. *Numerical mathematics*, volume 37. Springer Science & Business Media, 2010.